Three essays on market frictions

Dissertation
submitted to the Faculty of Economics, Business Administration and Information Technology of the University of Zurich

to obtain the degree of
Doktorin der Wirtschaftswissenschaften, Dr. oec.
(corresponds to Doctor of Philosophy, PhD)

presented by
Lujing Su
from China

approved in April 2015 at the request of

Prof. Dr. Markus Leippold
Prof. Dr. Alexandre Ziegler

The Faculty of Economics, Business Administration and Information Technology of the University of Zurich hereby authorizes the printing of this dissertation, without indicating an opinion of the views expressed in the work.

Zurich, 01.04.2015.

The Chairman of the Doctoral Board: Prof. Dr. Josef Zweimüller.

Acknowledgments

In the process of completing this thesis, I have obtained a lot of feedback and support for which I am very grateful and feel deeply indebted. First and foremost, I would like to thank my supervisor Prof. Dr. Markus Leippold, who has guided me and pointed me in the right direction throughout the years that I worked on my thesis. Without his commitment and scholarly guidance, this dissertation would have never materialised as it is today.

I am also extremely indebted and grateful to my co-supervisor Prof. Dr. Alexandre Ziegler for his support during my Ph.D.; whenever I needed help, he was there to share his knowledge and to guide me. His invaluable advice on all aspects of the thesis and deep knowledge helped a lot in improving the quality of the thesis.

I am also grateful to Prof. Dr. Loriano Mancini for attending my thesis defense and Prof. Dr. Marc Paolella for always being encouraging and supportive since the start of the Ph.D. program.

Moreover, I would like to highlight that I enjoyed the company, discussions, and support of many colleagues and friends at the Department of Banking and Finance including Elise Gourier, Chris Bardgett, Felix Matthys, Nikola Vasiljevic, Meriton Ibraimi, Felix Stang, Zexi Wang, Nilufer Caliskan, Jakub Rojcek, Robert Huitema, Quan Zhang, Elisabeth Megally, Pascal Marco Caversaccio and many others. I am also especially grateful to Tatjana-Xenia Puhan for her continuous support and suggestions, and also for our jogging dates on Thursday evenings. I would also like to thank Dörte Kreher and Manish Gupta for many wonderful dinners that we had together and their friendship that warms my heart.

I would also like to express my warmest thanks to my parents, my husband Hao Wei, my flatmate Yijian Gong and many other friends for their support along a profoundly interesting but difficult journey. I may have started this, but I certainly would have struggled to finish it without you.

Contents

List of Tables

List of Figures

Chapter 1

Preface

Standard financial market models are based on the simplifying assumption that trading can be performed without frictions (i.e. zero transaction costs and no institutional restrictions on trading). Such assumption is the building block of the capital asset pricing model (CAPM) and its extensions, such as the multi-period ICAPM and the consumption CAPM. The assumption of frictionless markets also underlies derivative pricing models such as the Black-Scholes model in which the absence of arbitrage opportunities allows obtaining the derivative price by setting up a perfect replicating portfolio using the underlying and riskless bonds. The assumption of a frictionless market is clearly a strong idealization of reality, since ample empirical evidence demonstrates the existence of sizable market frictions.

The two most important types of market frictions are transaction costs and institutional restrictions on trading. Transaction costs can either be interpreted as the explicit costs charged on trades in the form of commission fees and taxes or other implicit costs such as the bid-ask spread. Institutional restrictions take the form either of prohibitions on particular classes of trades, or of conditions that must be fulfilled before trades are permitted. One type of institutional restrictions on trading that has attracted relatively little attention from researchers are margin requirements. Margin requirements that are often imposed by exchanges not only restrict the amount of money investors can borrow, but also specify the minimum amount of cash (equity) investors have to deposit with the exchange to cover potential losses. Besides these two types of market frictions, other types of frictions include asymmetric information and differences between lending and borrowing rates.

There are many attempts in the literature to build asset and derivative pricing models incorporating transaction costs and trading constraints. These frictions have been found to have a sizable impact on the trading behavior of market participants, optimal asset allocations and equilibrium asset prices and hence alter many of the conclusions of traditional theory. This thesis contributes to the literature by investigating the impact of various market frictions on asset prices (both stocks and options) and on information and portfolio choice.

This thesis mainly answers three questions. First, what is the impact of imposing margin requirements on option prices? Second, index futures and exchange-traded funds (ETFs) are index derivatives with payoffs very close to that of the index. However, when taking transaction costs into account, such index derivatives are not redundant, because they provide investors a cheap way to gain exposure to the broad market. Given the increasing popularity of index futures and ETFs among investors, what is the impact of trading activity in index derivatives on underlying stocks' correlations? Third, investors in financial markets have incomplete information regarding the payoff of the assets. How investors allocate their information has important implications for their portfolio choices. Do transaction costs, a proxy for liquidity, also play a role in determining information acquisition? Each essay of the thesis answers one of the above questions.

In the first chapter *collateral smile*, we analyze the impact of funding costs and margin requirements on index options traded on the CBOE. Assuming differential borrowing and lending rates, we derive no-arbitrage bounds for European options. We show that funding costs and the CBOE's margin requirements lead to an increase in option prices, which translates into skew and smile patterns for implied volatility curves even under constant volatility. Empirical tests confirm that our model-implied slopes have significant statistical power in explaining the slopes observed in the market. Hence, at least in part, funding costs and collateral requirements provide an institutional explanation of the volatility smile phenomenon.

In the second chapter *how index futures and ETFs increase stock return correlations*, we examine whether increased trading activity in index futures and exchange traded funds (ETFs) is associated with higher equity return correlations. We build a simple model to analyze how demand shocks for index ETFs and futures are transmitted to the underlying stocks through arbitrage. Our model predicts that demand shocks to ETFs and futures lead to a stronger price comovement not only for index stocks but also for non-index stocks. Moreover, demand shocks to physical (rather than synthetic) index ETFs have a higher impact on stock return correlations than demand shocks to futures. We confirm the model predictions by studying the average pairwise correlation of S&P 500 stocks after the inception of S&P 500 futures. Controlling for several factors, we show that trading activity in futures and ETFs explains the time variation of the average S&P 500 stock return correlation with ETFs exhibiting a significantly stronger explanatory power. An examination of the relationship between current and lagged returns suggests that at least some of the return comovement is excessive. Furthermore, we confirm that trading activity in futures and ETFs is also associated with higher return correlations among non-index stocks.

In the third chapter *the impact of liquidity on information acquisition*, we extend the information acquisition problem considered in Van Nieuwerburgh and Veldkamp (2010) to a more realistic setting with transaction costs, including both proportional and quadratic transaction costs. Our findings differ from those obtained without transaction costs. As an asset's transaction costs rise, it becomes less attractive for investors to learn about it. Investors' decision about which assets to learn depends on their initial holdings. Moreover, transaction costs might change investors' information acquisition policy from specialized learning to generalized learning. Hence, in addition to the assets' Sharpe ratio, investors' initial asset holdings and assets' liquidity play an important role in determining optimal information acquisition strategies.

Chapter 2

Collateral Smile

with Markus Leippold

This paper is accepted for publication in Journal of Banking and Finance, 2015, Volume 58, Page 15-28, Copyright by Elsevier B.V.

Abstract

We analyze the impact of funding costs and margin requirements on index options traded on the CBOE. Assuming differential borrowing and lending rates, we derive no-arbitrage bounds for European options. We show that funding costs and the CBOE's margin requirements lead to an increase in option prices, which translates into skew and smile patterns for implied volatility curves even under constant volatilities. Empirical tests confirm that our model-implied slopes have significant statistical power in explaining the slopes observed in the market. Hence, at least in part, funding costs and collateral requirements offer an institutional explanation of the volatility smile phenomenon.

We thank Panayiotis Andreou, David Bates, Jonathan Berk, Tony Berrada, Josh Coval, Stephen Figlewski, Rajna Gibson, Lasse H. Pedersen, Adriano Rampini, Paul Whelan, Alexandre Ziegler and the participants at the Revelstoke Finance Seminar, the Conference of the Swiss Society for Financial Market Research, the 18th Annual Conference of the Multinational Finance Society, Swiss Finance Institute Asset Pricing Workshop 2011, AFA 2013 San Diego, and the finance seminars at the University of Geneva and the University of Zurich for useful comments. Financial support by the Swiss Finance Institute (SFI), Bank Vontobel, and the National Center of Competence in Research "Financial Valuation and Risk Management" is gratefully acknowledged.

2.1 Introduction

We analyze the impact of funding costs and margin requirements on the prices of index options traded on the Chicago Board Options Exchange (CBOE). Margin requirements are collateral the option sellers are required to deposit with the exchange. Funding costs refer to the spread between borrowing and lending rates. We propose a model that gives upper and lower bounds for option prices in the absence of arbitrage in a dynamically incomplete market with differential borrowing and lending rates. We shows that funding costs and margin requirements generate arbitrage bounds that allow for skew and smile patterns for implied volatilities (IV) that are consistent with what we typically observe in option markets. Empirical tests show that our model-implied slopes have significant statistical power in explaining the slopes observed in the market.

Imposing margin requirements or collateral requirements is common practice in both over the counter (OTC) and exchange-based transactions. During the recent financial crisis, market participants had to painfully acknowledge that the value of a derivative depends not only on its payoff structure, but also on the counterparty's creditworthiness. To mitigate counterparty risk, the contracting party for whom the derivative has a negative value is required to deposit collateral on a margin account to guarantee a certain recovery rate in case of default. While on OTC markets the use of collateral became widespread only over the past few years, standardized margin requirements have been used at exchanges already since the late 1980s. Yet, for the most part, the option pricing literature has been silent on how these margin requirements influence the pricing of exchange-traded derivatives. Therefore, we take a closer look at the price impact of collateral rules on exchange-traded index options.

A critical quantity for our analysis is the spread between borrowing and lending rates, which may become particularly large during financial crises. This spread captures the difference between the benefit and cost of depositing collateral. The benefit is the interest rate the investor receives from the entity where the collateral is deposited. This rate is usually equivalent to the lending rate. The cost of the collateral refers to the interest rate the investor must pay on the collateral amount if borrowed from another entity. The wedge between the benefit and cost of posting collateral is the channel through which collateral requirements affect derivative prices.

For the pricing of options, the funding costs, measured as the spread between borrowing and lending rates, must also influence the replicating strategy. The money needed to purchase the underlying and to be deposited in the margin account is borrowed at a rate that exceeds the rate at which short selling proceeds can be invested. Indeed, assuming a significantly higher borrowing rate is in line with the currently prevailing market conditions. In the recent financial crisis, banks had difficulties in funding and maintaining a certain level of liquidity. These difficulties were further exacerbated by mutual distrust and an increasing reluctance to lend money to one another. Very quickly, interbank money markets dried out. In particular, cash lending became quite restricted and other key funding sources were also inaccessible.

A commonly agreed-upon measure of funding difficulty is the Libor–OIS spread, defined as the difference between the interest rates on interbank loans and the Overnight Index Swap rate. Between 2002 and the beginning of the recent financial crisis, the three-month Libor–OIS spread was usually around 10 to 30 basis points (bps). However, it jumped to 66 bps on August 20, 2007, and remained high until March 2009, with a peak of 364 bps on October 10, 2008. In a situation in which the historically stable Libor–OIS spread varies dramatically and rises to new levels, the assumption of a single risk-free rate for borrowing and lending is no longer appropriate. The wedge between these two rates in interplay with collateral requirements may then have an economically significant impact on the pricing of derivatives.

Motivated by the increased importance of collateralization in the aftermath of the 2008 financial crisis, we put forward a model which takes these two market frictions into account. To isolate the effect of collateral requirements and funding costs on option prices, we choose the classical Black–Scholes model as our starting point. However, we work in an incomplete market framework as in Bergman (1995), which allows us to drive a wedge between the borrowing and lending rate. In an incomplete market, a unique equilibrium option price can only be derived when additional assumptions on the structure of the economy are made. Nevertheless, the absence of arbitrage allows us to put meaningful bounds on option prices. Hence, we extend the model of Bergman (1995) by incorporating collateral requirements and we derive solutions for the upper and lower bounds of option prices. We find that the lower bound is equal to the price given by the standard Black–Scholes formula with the proper interest rate inserted. However, the upper bound depends on both borrowing and lending rates as well as the specification of the collateral requirements. Furthermore, we can decompose the resulting upper bound for option prices into the traditional Black–Scholes price and an additional margin adjustment part.

Depending on the margin rules, the exact form of the option upper prices varies for different exchanges. We investigate explicitly the impact of the margin requirements imposed by the CBOE. By choosing parameter values based on historical data, we show that this margin adjustment plays a non-negligible role in determining upper bounds of option prices. Furthermore, its relative importance varies with moneyness. We illustrate numerically that the option IV bounds accounting for margin requirements and funding costs as imposed by the CBOE are capable of allowing for substantial volatility smiles, similar in magnitude to those observed in the data. This feature of our model does not rely on jumps or stochastic volatilities of the underlying price processes, which may already and in part explain the observed volatility smile. Hence, not only deviations from the geometric Brownian motion assumption, such as jumps and stochastic volatility, but also the general institutional set-up of the market may be responsible for a significant part of the observed IV patterns.

To investigate whether the above claim also holds under more general assumptions regarding the underlying's stochastic process, we extend the model to allow for stochastic volatility as in Heston (1993). However, introducing stochastic volatility requires additional assumptions on the replicating strategy. We find that, also in the presence of stochastic volatility, the upper bounds of the IV taking into account collateral requirements and funding costs show a significant increase from the IV as implied by Heston's model. Qualitatively, the impact is the same as in the constant volatility case.

Bringing our model to the data seems to be a promising next step. In particular, we challenge our constant volatility model by testing whether we could generate volatility slopes comparable with the empirical ones. Following the methodology applied in Bakshi, Kapadia, and Madan (2003) (BKM hereafter), we find a clear link between the empirical slope and the slope predicted by our model. A simple ordinary least square regression (OLS) on the differences shows that, on average, our theoretical slope changes can already account for more than 30% of the time variation of the empirical slope changes. Therefore, our model provides an additional avenue to explain at least in part the variation of IV smiles.

Taking margin and funding costs into account is not completely new in derivative pricing. For instance, Johannes and Sundaresan (2007) discuss the impact of collateral on swap prices. Using Eurodollar futures rates, they found that swaps are priced above the traditional portfolio of forwards value and below a portfolio of futures value. Berkovich and Shachmurove (2013) argue that the collateral requirement for a trading strategy is path dependent. Once the actual cost of implementing a put selling strategy is fully taken into account, writing put options on S&P 500 index (SPX) earns only normal returns or even negative returns. Lou (2009) shows how the recently observed negative swap spread can be explained by asymmetric funding costs.

Our study is also related to papers that investigate option pricing bounds when the Black–Scholes assumption of a dynamically complete and frictionless market is violated. In an incomplete market, the usual replication argument is not applicable, because there are not enough basis assets to span the uncertainty. In the presence of market frictions such as, e.g., short selling constraints and transaction costs, the no-arbitrage argument alone is not enough to determine a unique option price. Instead, option prices must lie in a band that corresponds to the expected value of the option payoff under all the measures that rule out arbitrage. To determine these bounds, one approach focuses on finding the minimum costs to hedge (see, e.g., Cvitanic, Pham, and Touzi (1998) and Cvitanic, Pham, and Touzi (1999)). Another approach obtains tighter bounds by eliminating stochastically dominating strategies in comparing two portfolios by assuming a risk-averse investor (see, e.g., Perrakis and Ryan (1984), Levy (1985), and Ritchken (1985)). A third approach tightens the bounds by imposing assumptions on the pricing kernel, such as its volatility or on the gain–loss ratio (see, e.g., Cochrane and Saa-Requejo (2000) and Bernardo and Ledoit (2000)). In our model, we follow the first approach by restricting the equilibrium price of options to a band where arbitrage opportunities are ruled out.

The papers closely related to our study are Santa-Clara and Saretto (2009), Bergman (1995) and Piterbarg (2010). However, our work differs from these papers in at least four ways. Firstly, Santa-Clara and Saretto (2009) argue that the margin calls and funding costs could make the strategy involving selling OTM puts unprofitable, and thus OTM put options remain expensive. Our paper, however, studies directly the impact of these two market frictions on option prices.

Secondly, Bergman (1995) studies the impact of funding costs on option prices and derives the resulting no-arbitrage bounds. However, he does not consider the impact of collateral requirements at all, which may lead to some counterintuitive results when inverting the no-arbitrage bounds for prices to no-arbitrage bounds for IVs. In particular, in the model of Bergman (1995) the upper no-arbitrage bound for put options degenerates to a constant. Hence, the existence of differential borrowing and lending rates cannot generate any smile pattern.

Thirdly, allowing for differential borrowing and lending rates is a complication that Piterbarg (2010) does not consider. Piterbarg (2010) introduces the intricacy of differential rates based on the types of assets that are used to secure the funding, but the same rate is used for borrowing and lending. In contrast, our paper looks at the impact of differential borrowing and lending rates on option prices. In addition, in Piterbarg (2010) the probability measure is implicitly fixed without further specification. Hence, there are unique option prices. However, in our model the analysis is based on no-arbitrage bounds, since the market is inherently imperfect due to the wedge between borrowing and lending rate.

Fourthly, we provide evidence on the actual impact on option prices of the collateral rules as explicitly specified by the CBOE. Furthermore, using option price data, we also test the performance of our model by fitting empirical IV curves. To our best knowledge, these aspects have not been considered by previous papers.

We organize the paper as follows. In Section 2.2, we provide an overview of the collateral requirements for options traded on the CBOE. In Section 2.3, we develop an option pricing model that accounts for funding costs and margin requirements. We derive the upper and lower bounds of option prices under the CBOE margin rules. In Section 2.4, we analyze the margin-based model and the resulting IV curves numerically. Section 2.5 extends the model to allow for stochastic volatility. In Section 2.6, we bring our model to the data and conduct an empirical study of IV slopes. Section 2.7 concludes.

2.2 Margin Requirements for Derivatives in Practice

Collateral requirements on exchanges, usually referred to as margin requirements, are set by each exchange individually, and may differ across markets. For our analysis, we restrict ourselves to the world's largest option trading exchange, the CBOE. We explicitly focus on the margin requirements for the index options traded on the CBOE. In what follows, we briefly explain these margin requirements and we refer the interested reader to the CBOE's website for detailed explanations and specific examples.[1]

Margin requirements for buyers and sellers of options differ. Option buyers, who obtain a right rather than an obligation, are exempted from margin requirements once the full price of the option is paid. The reason is simple: buyers can always let the option expire without incurring further costs. Moreover, on the CBOE, for options with a time to expiration of more than nine months, buyers are allowed to pay 75% of the cost of the options as the initial margin with a maintenance margin at 75% of the option market value. In the following analysis, we assume that buyers pay the option price in full, since most liquid options have short maturities and, hence, need to be paid in full.

Writers of options are required to post margins to cover the risk of no delivery (when asked) at maturity. For example, writing a call option generates the risk of an unlimited loss, as the underlying price can increase to an arbitrarily large value. Therefore, call option sellers are required to deposit cash in the margin account to protect buyers against the sellers' default. The use of clearing houses guarantees that the option contract will be fulfilled. Therefore, we do not take into account the option writer's default risk in our model.

For option sellers, the CBOE uses two alternative margin rules, the strategy-based margin rules and the portfolio margining rules. Under the strategy-based margin rules, the positions are managed under the so-called strategy margin account and the margin is calculated according to each predefined option strategy.[2] Strategy-based margin rules have been effective since the 1980s. In a private communication from the CBOE, we were informed that the strategy-based margin rules still remain effective for a significant part of the options traded on the CBOE. Therefore, we include these rules in our analysis.

Strategy-based margin rules use predefined formulas to compute margin requirements based on the strategy option writers apply. For a naked option traded on the CBOE, the strategy-based margin rule consists of the option market value and some portion of the underlying value or strike price, and is

$$\begin{aligned} \text{Call:} \quad & C(t) = \max(V(t) + a_1 S(t) - (K - S(t))^+, V(t) + a_2 S(t)), \\ \text{Put:} \quad & C(t) = \max(V(t) + a_1 S(t) - (S(t) - K)^+, V(t) + a_2 K), \end{aligned} \tag{2.1}$$

where $C(t)$ is the margin amount, $S(t)$ is the underlying price, $V(t)$ is the value of the option, and the parameters a_1 and a_2 represent the margin parameters specified by the CBOE. For options on a broad index, the CBOE currently sets the parameters a_1 and a_2 equal to 0.15 and 0.1, respectively. For equity options or options written on a narrow based index, a_1 and a_2 are set equal to 0.2 and 0.15. Note that these are the minimal margin requirements for strategy-based margin accounts for all types of investors, including brokers. Individual investors are

[1] The CBOE margin manual can be downloaded from http://www.cboe.com/tradtool/marginmanual2000.pdf. It provides a complete description of the margin requirements for the various option strategies.

[2] Examples of such strategies are, e.g., a short put, covered call, long vertical call spread, etc.. The CBOE provides a margin manual on its website to explain the details of the margin requirements for each type of strategy.

sometimes subject to much higher margin parameters charged by the brokerage firms, which could reach 40% for a_1 and 35% for a_2 (Santa-Clara and Saretto (2009)).

On April 2, 2007, the CBOE amended the margin rules and introduced the portfolio margining rules, which allow charging margins based on the risk exposure of the whole option portfolio. For some positions, the margin requirements may not have changed significantly, but for positions with offsetting exposures, the difference can indeed be significant. The portfolio margining rule is a scenario-based rule that calculates the possible losses assuming various market moves. For SPX related products, the market moves in the underlying index are specified within a range of -8% to +6%. The computed largest potential loss must then be compared with a per contract minimum of 37.50 dollars (for SPX options with multiplier 100). The greater of these two defines the margin requirement. Currently, the option pricing model that the CBOE uses for computing the possible loss for option positions upon various market moves is not publicly available. Hence, the best thing we can do for our numerical analysis, is to assume that the CBOE uses the standard Black-Scholes formula to determine portfolio margins.

We consider two types of portfolio margin requirements, namely the margin requirement for a naked short sale and the minimum margin requirement. The naked short sale portfolio requirement assumes there is only one option in the portfolio margin account, while the minimum portfolio margin requirement considers the least amount of capital that must be locked in the account for every option sold. Margin requirement is the amount of equity (cash) that must be maintained in a margin account. It is calculated as the sum of the market value of all long positions minus the sum of the market value of all short positions. Note that whenever an option is written in the portfolio, the cash balance generated by selling the option is usually kept in the account to offset the short position created by option writing. The margin for each option is therefore larger than the value of the option. The naked short-sell margin requirement under portfolio margin account, to be more specific, is

$$C(t) = \max_{k\in K}\{V((1+k)S(t),t), V(t)+37.50\}, \tag{2.2}$$

where $K = \{-0.08, -0.07, ..., 0.05, 0.06\}$ is the set of market scenarios and $V((1+k)S(t),t)$ denotes the option price when the underlying moves from $S(t)$ to $(1+k)S(t)$. When the underlying value moves to $(1+k)S(t)$, the loss generated from writing the option is $V((1+k)S(t),t) - V(t)$. Adding up with the proceeds from option writing $V(t)$ yields $V((1+k)S(t),t)$. The margin requirement is thus the greater of the worst possible loss and the per contract minimum. For a naked short sale, as there is only one option in the account, it is straightforward that $C(t) = \max\{V((1.06S(t),t), V(t)+37.50\}$ for calls and $C(t) = \max\{V(0.92S(t),t), V(t)+37.50\}$ for puts.

Realistically, investors hold not only one, but many options in their trading account. Hence, we must also analyze the margin requirement for writing an option when the investor is holding already a portfolio involving many options. This margin requirement depends on the loss

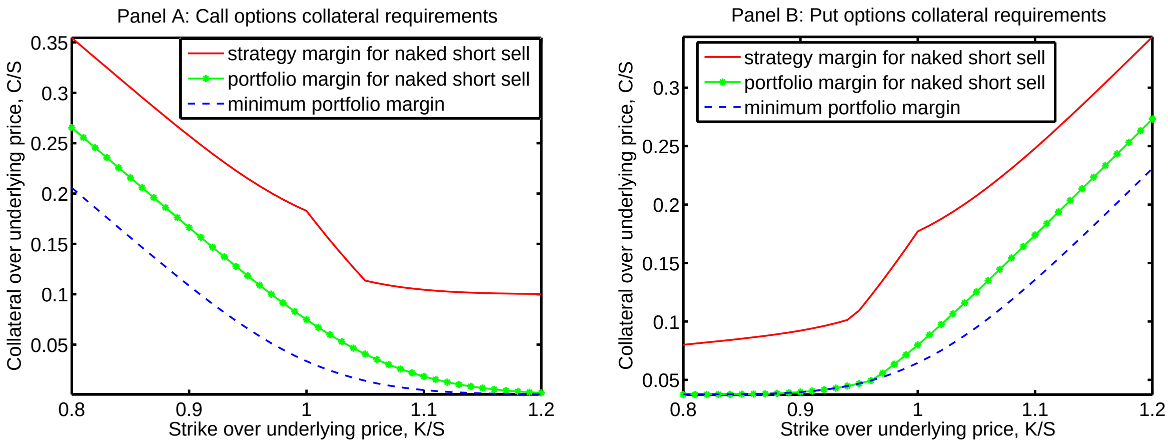

FIGURE 2.1: *Margin requirements for put and call options on the CBOE.*
The figure plots the margin requirements as imposed by the CBOE as a function of moneyness. Panel A plots the strategy-based margin requirements for short selling calls, the portfolio margin requirements for short selling calls and the minimum portfolio margin requirements for selling calls. Panel B plots the strategy-based margin requirements for short selling puts, the portfolio margin requirements for short selling puts and the minimum portfolio margin requirements for selling puts. Margin requirements are computed by assuming the option prices are given by the Black–Scholes formula for a maturity of three months. Margin requirements for other maturities and other models are similar in magnitude and share the same qualitative features.

and profit on the composition of the corresponding portfolio. Due to the lack of data on typical portfolios held at the CBOE, we circumvent this problem as follows. Instead of considering arbitrary portfolio compositions, we consider only the minimum margin requirement that writing an option incurs. In particular, the least possible margin requirement for a short option position in the portfolio is simply the sum of the per contract minimum and the option proceeds, i.e., $V(t)+37.50$. Note that this minimum portfolio margin is the least possible margin for any type of strategy.[3] By using this minimum requirement, we get a conservative estimate of the margin's impact under the portfolio margining rules.

In the subsequent analysis, we use three types of margining rules, the strategy-based margin rules for a naked short sale, the portfolio margining rule for a naked short sale, and the minimum possible portfolio margining rule for writing an option. In Figure 2.1, we illustrate these three types of margin rules for various moneyness levels for a call option (Panel A) and a put option (Panel B). We see that the margin requirements are the highest for ITM options and then gradually decrease when options become OTM. Among all the margin rules we consider, the strategy-based margin requirement for a short sale is the most stringent, while the minimum portfolio margining requirement is the least.

[3]For example, for a covered call strategy, the option seller also needs to satisfy the minimum portfolio requirement.

2.3 Option Pricing with Costly Margin Requirements

In terms of the price dynamics, we base our analysis on the standard Black–Scholes assumptions. However, we depart from the Black–Scholes model by introducing differential borrowing and lending rates as well as margin requirements for option writers. As Bergman (1995) argues, a dynamically incomplete capital market allows the existence of a wedge between borrowing and lending rates. Depending on the structure of the market, equilibrium option prices may depend on other state variables. Even though a pure no-arbitrage argument cannot uniquely determine option prices, we can derive option pricing bounds, the violation of which indicates arbitrage opportunities even after accounting for market imperfections.

To analyze the option pricing problem with differential borrowing and lending rates, we introduce three accounts. The first is a cash account, where cash is deposited to finance the purchase of the underlying and to hold the proceeds from short selling the underlying. It plays the role of a traditional savings account where the deposited cash earns the lending rate r_l and borrowing is not allowed. Our second account is a debt account, from which the option writer can borrow the funds used for the replicating portfolio if the writer's cash holding is not enough. The debt account is charged at the borrowing rate r_b. The third account is the collateral account to secure the margin requirement. The deposited cash earns the lending rate r_l. To simplify the computations, we assume the borrowing and lending rates are constant. In general, we have $r_b \geq r_l$, as the spread between the two rates reflects the return the bank must earn for its operations.

2.3.1 No-arbitrage Bounds

Within our incomplete market setting, we cannot derive a unique option price unless we impose some additional structure. However, arbitrage considerations help us to derive pricing bounds on the options. To obtain these bounds, we need to analyze the portfolio strategy that replicates the payoff of the option at expiration. The replicating strategy in our case is defined by a four-dimensional process $(\alpha(t), \beta(t), \lambda(t), \delta(t))$ to capture the different interest rates earned on different accounts. By $\alpha(t)$, we denote the amount of stocks that we hold at time t; by $\beta(t) < 0$, the cash borrowed from the debt account; by $\lambda(t) > 0$, the cash deposited at the cash account; and by $\delta(t)$, the cash deposited in the collateral account.

To prevent arbitrage, we can show that the option price $V(t)$ with a payoff of $h(S(T))$ at expiration $T \geq t \geq 0$ must lie within an upper and a lower bound. The underlying price is denoted by $S(t)$ with a continuous dividend yield r_d. We first focus on the lower bound, and consider the following minimization problem:

$$\mathcal{M}^{-} \quad : \quad \min_{\alpha(t),\beta(t),\lambda(t),\delta(t)} V(0), \tag{2.3}$$

subject to

$$
\begin{aligned}
V(t) &= \alpha(t)S(t) + \beta(t) + \lambda(t) + \delta(t), \\
dV(t) &= \alpha(t)(dS(t) + r_d S(t)dt) + r_b\beta(t)dt + r_l\lambda(t)dt + r_l\delta(t)dt, \\
V(T) &\geq -h(S(T)), \\
\delta(t) &\geq C(t, S(t)) \text{ for option buyers.}
\end{aligned}
$$

We denote the solution to the $\mathscr{M}^-$-problem by V_0^-. Note that an investment used to replicate a non-positive payoff must have a non-positive initial capital, hence V_0^- is less than or equal to zero. Obviously, $V \geq -V_0^-$ has to hold, otherwise there is an arbitrage opportunity. We could buy the option and implement the strategy that solves the $\mathscr{M}^-$-problem. This strategy meets the collateral requirement for the option buyer and gives a payoff that is greater than $-h(S(T))$. The combined payoff thus gives a non-negative payoff at maturity and generates a positive cashflow at the initial time, $-V_0^- - V > 0$. Since the collateral requirements for option buyers are zero, as discussed in Section 2.2, $\delta(t)$ is zero in the optimal solution. Therefore, collateral does not play a role in determining V_0^-.

To determine the upper arbitrage bounds, we consider the following optimization problem,

$$
\mathscr{M}^+ \quad : \quad \min_{\alpha(t),\beta(t),\lambda(t),\delta(t)} V(0), \tag{2.4}
$$

subject to

$$
\begin{aligned}
V(t) &= \alpha(t)S(t) + \beta(t) + \lambda(t) + \delta(t), \\
dV(t) &= \alpha(t)(dS(t) + r_d S(t)dt) + r_b\beta(t)dt + r_l\lambda(t)dt + r_l\delta(t)dt, \\
V(T) &\geq +h(S(T)), \\
\delta(t) &\geq C(t, S(t)) \text{ for option writers.}
\end{aligned}
$$

We denote the solution to the $\mathscr{M}^+$-problem by V_0^+. $V \leq V_0^+$ has to hold if arbitrage opportunities are to be excluded. When $V > V_0^+$, selling the option and employing the strategy that solves the $\mathscr{M}^+$-problem is an arbitrage opportunity. This strategy satisfies the collateral requirement for option writers and gives a payoff greater than $+h(S(T))$. Therefore, the combining strategy has a non-negative payoff at maturity and generates a positive cashflow at initiation, i.e., $V - V_0^+ > 0$. We can now summarize the above discussion in the following proposition.

Proposition 1 *In a dynamically incomplete market with $r_l \neq r_b$ and with collateral requirements, the option price V_0^e must lie within the arbitrage band $[-V_0^-, V_0^+]$, where V_0^- and V_0^+ solve $\mathscr{M}^-$ and $\mathscr{M}^+$, respectively.*

As collateral has no impact on the lower bound of option prices, the lower bound corresponds exactly to the one derived by Bergman (1995), who also considers differential borrowing and lending rates. Therefore, we borrow the following result:[4]

Proposition 2 (Bergman (1995)) *In the Black–Scholes setting, but under differential borrowing and lending rates, the lower bound for calls is given by the classical Black–Scholes call option formula with the lending rate replacing the risk-free rate. For put options, the risk-free rate is replaced by the borrowing rate.*

However, for the determination of the upper bounds, i.e., the solution to $\mathscr{M}^+$, we cannot rely on Bergman (1995), as he does not take collateral into account.

2.3.2 General Formulas for Upper Price Bounds

In the Black–Scholes model, borrowing or lending occurs at the same interest rate. Therefore, the same PDE applies for the pricing of both puts and calls, but with different boundary conditions. However, in the presence of funding costs, the replicating strategy for puts and calls involves different positions in the cash, debt, and collateral accounts. This leads to subtle differences in the PDE representation of calls and puts. In the case of a call option, we must carefully segregate the positions into *i)* a collateral $C(t)$ required by the exchange to be deposited in the cash account earning the lending rate, *ii)* the quantity $V(t) - C(t)$ borrowed at the borrowing rate from the debt account to finance the posting of margin, and finally *iii)* $\alpha(t)S(t)$ borrowed from the debt account to finance the stock purchase.[5] In the case of a put option, we have to track separately the positions in the cash, debt, and collateral accounts by decomposing them as above into *i)* the collateral $C(t)$ deposited in the cash account, *ii)* the quantity $V(t) - C(t)$ borrowed to finance the required margin, and *iii)* the short selling proceeds $\alpha(t)S(t)$ deposited in the cash account.[6] We summarize the resulting pricing formulas below. The proof is given in Appendix 2.A.

Proposition 3 *In the Black–Scholes setting, but under differential borrowing and lending rates, the upper bound for call options in the presence of collateral requirements is given by the expectation*

$$V_{call}(t) = \mathbb{E}_t^{\mathbb{P}_b}\left[e^{-r_b(T-t)}V(T) + \int_t^T e^{-r_b(u-t)}(r_b - r_l)C(u)du\right] \tag{2.5}$$

under the pricing measure $\mathbb{P}_b$, subject to $V_{call}(T) = (S(T) - K)^+$ and

$$dS(t)/S(t) = (r_b - r_d)dt + \sigma dW^b(t),$$

[4] We do not repeat the derivation here, but refer to Bergman (1995) for details.

[5] Even under the portfolio-based margin rule, the proceeds of selling options must be kept in the margin account. Therefore $V(t) - C(t)$ is indeed borrowing.

[6] When short selling stocks, the proceeds are usually kept with the broker and cannot be used by the investor.

where $W^b(t)$ is a standard Brownian motion under $\mathbb{P}^b$. The corresponding upper bound for put options is given by the expectation

$$V_{put}(t) = \mathbb{E}_t^{\mathbb{P}_l}\left[e^{-r_b(T-t)}V(T) + \int_t^T e^{-r_b(u-t)}(r_b - r_l)C(u)du\right] \tag{2.6}$$

under the pricing measure $\mathbb{P}_l$, subject to $V_{put}(T) = (K - S(T))^+$ and

$$dS(t)/S(t) = (r_l - r_d)dt + \sigma dW^l(t),$$

where $W^l(t)$ is a standard Brownian motion under $\mathbb{P}_l$.

Intuitively, the pricing formulas in the proposition have two components. For instance, in the case of the call option, the first component $\mathbb{E}_t^{\mathbb{P}_b}\left[e^{-r_b(T-t)}V(T)\right]$ plays the role of the traditional Black–Scholes call option price, but with a different probability measure and discount rate. The second part, $\mathbb{E}_t^{\mathbb{P}_b}\left[\int_t^T e^{-r_b(u-t)}(r_b - r_l)C(u)du\right]$, reflects the impact of the margin requirements on the option price, and we refer to it as the margin adjustment. Since $C(t) > 0$, the margin adjustment is always positive. We can interpret it as the additional price the writer requires to be compensated for the increasing replication cost induced by fulfilling the margin requirements. If it is costless to post collateral, i.e., if the collateral earns the same rate as the borrowing rate $r_b = r_l$, then the margin adjustment disappears and the margin requirement would not influence the call price at all. Indeed, when $r_b = r_l$, equations (2.5) and (2.6) collapse to the standard Black–Scholes formula. However, whenever $r_b > r_l$, which is usually the case, the margin requirements increase the replicating cost and the call option prices through the margin adjustment.

It is worth noting that Proposition 3 provides a general formula to compute upper bounds on option prices under margin constraints and funding costs. Even though we focus on SPX options traded on the CBOE, its application is not restricted to this particular case.

2.3.3 Upper Price Bounds under the CBOE's Margin Requirement

Having derived the general option pricing formula in the presence of funding costs and general margin requirements, we can now insert the specific margin rule of the CBOE into the pricing formula to obtain the upper bound under the actual margin rules. We consider for our analysis three margin requirements: the strategy margin requirement for a naked short sale, the portfolio margin requirement for a naked short sale, and the minimum portfolio margin requirement. We collect the corresponding formulas in the corollaries below, which follow directly from Proposition 3 and are proven in Appendix 2.B.

For options subject to the CBOE's strategy margin requirements, the upper bound for a short sale can be derived closed form as shown in the following corollary.

Corollary 1 *At time t, the upper price bound for European call options subject to CBOE's strategy margin rules for a short sale with maturity T, strike K is*

$$\begin{aligned}
V_{call}(t) &= S(t)e^{(r_b-r_d-r_l)(T-t)}N(d_1(T,t)) - Ke^{-r_l(T-t)}N(d_2(T,t)) \\
&\quad +(r_b-r_l)S(t)\int_t^T e^{(r_b-r_d-r_l)(u-t)}\left(a_2 N(-d_1^*(u,t)) + a_1 N(d_1(u,t))\right)du \\
&\quad +(1+a_1)(r_b-r_l)S(t)\int_t^T e^{(r_b-r_d-r_l)(u-t)}\left(N(d_1^*(u,t)) - N(d_1(u,t))\right)du \\
&\quad -(r_b-r_l)K\int_t^T e^{-r_l(u-t)}\left(N(d_2^*(u,t)) - N(d_2(u,t))\right)du,
\end{aligned}$$

with

$$\begin{aligned}
d_{1,2}(u,t) &= \frac{\ln(\frac{S(t)}{K}) + (r_b - r_d \pm \frac{1}{2}\sigma^2)(u-t)}{\sigma\sqrt{u-t}}, \\
d_{1,2}^*(u,t) &= \frac{\ln(\frac{S(t)(1+a_1-a_2)}{K}) + (r_b - r_d \pm \frac{1}{2}\sigma^2)(u-t)}{\sigma\sqrt{u-t}},
\end{aligned}$$

where $N(\cdot)$ denotes the standard normal cumulative distribution function.

We remark that these pricing formulas are somewhat lengthy, but are merely the sum of the classical Black–Scholes price and the margin adjustment term. Analogously, we can derive the upper bound for the put option value.

Corollary 2 *At time t, the upper price bound for European put options subject to CBOE's strategy margin rules for a naked short sale with maturity T, strike K is*

$$\begin{aligned}
V_{put}(t) &= Ke^{-r_l(T-t)}N(-d_2(T,t)) - S(t)e^{-r_d(T-t)}N(-d_1(T,t)) \\
&\quad +a_2(r_b-r_l)K\int_t^T e^{-r_l(u-t)}\left(N(-d_2^*(u,t)) + N(d_2^{**}(u,t))\right)du \\
&\quad +a_1(r_b-r_l)S(t)\int_t^T e^{-r_d(u-t)}\left(N(-d_1(u,t)) - N(-d_1^*(u,t))\right)du \\
&\quad +(r_b-r_l)K\int_t^T e^{-r_l(u-t)}\left(N(-d_2^{**}(u,t)) - N(-d_2(u,t))\right)du \\
&\quad +(a_1-1)(r_b-r_l)S(t)\int_t^T e^{-r_d(u-t)}\left(N(-d_1^{**}(u,t)) - N(-d_1(u,t))\right)du,
\end{aligned}$$

with

$$
\begin{aligned}
d_{1,2}(u,t) &= \frac{\ln(\frac{S(t)}{K}) + (r_l - r_d \pm \frac{1}{2}\sigma^2)(u-t)}{\sigma\sqrt{u-t}}, \\
d^*_{1,2}(u,t) &= \frac{\ln(\frac{a_1 S(t)}{a_2 K}) + (r_l - r_d \pm \frac{1}{2}\sigma^2)(u-t)}{\sigma\sqrt{u-t}}, \\
d^{**}_{1,2}(u,t) &= \frac{\ln(\frac{(1-a_1)S(t)}{(1-a_2)K}) + (r_l - r_d \pm \frac{1}{2}\sigma^2)(u-t)}{\sigma\sqrt{u-t}}.
\end{aligned}
$$

For the portfolio margin for a naked short sale, the margin requirements depend on the option pricing model, as the loss of the naked short sale is determined by the option value under various market moves. Therefore, we have to solve iteratively for the final option value by using standard numerical methods. For European call options, we get the following result.

Corollary 3 *At time t, the upper price bound for a European call option with maturity T, strike K, and subject to CBOE's portfolio margining rule for a naked short sale is*

$$
\begin{aligned}
V_{call}(t) &= S(t)e^{(r_b - r_d - r_l)(T-t)}N(d_1(T,t)) - Ke^{-r_l(T-t)}N(d_2(T,t)) \\
&\quad + \mathbb{E}_t^{\mathbb{P}_b}\left[\int_t^T e^{-r_l(u-t)}(r_b - r_l)(C(u) - V_{call}(u))du\right]
\end{aligned}
$$

with

$$
\begin{aligned}
d_{1,2}(u,t) &= \frac{\ln(\frac{S(t)}{K}) + (r_b - r_d \pm \frac{1}{2}\sigma^2)(u-t)}{\sigma\sqrt{u-t}}, \\
C(u) &= \max\{V_{call}(1.06S(u),u), V_{call}(u) + 37.50\},
\end{aligned}
$$

where $N(\cdot)$ denotes the standard normal cumulative distribution function.

Similarly, we can calculate the upper price bound for European put options under the portfolio margin rule for a naked short sale.

Corollary 4 *At time t, the upper price bound for a European put option with maturity T, strike K, and subject to CBOE's portfolio margining rule for a naked short sale is*

$$
\begin{aligned}
V_{put}(t) &= Ke^{-r_l(T-t)}N(-d_2(T,t)) - S(t)e^{-r_d(T-t)}N(-d_1(T,t)) \\
&\quad + \mathbb{E}_t^{\mathbb{P}_l}\left[\int_t^T e^{-r_l(u-t)}(r_b - r_l)(C(u) - V_{put}(u))du\right]
\end{aligned}
$$

with

$$\begin{aligned} d_{1,2}(u,t) &= \frac{\ln(\frac{S(t)}{K}) + (r_l - r_d \pm \frac{1}{2}\sigma^2)(u-t)}{\sigma\sqrt{u-t}}, \\ C(u) &= \max\{V_{put}(0.92S(u),u), V_{put}(u) + 37.50\}, \end{aligned}$$

For the minimum portfolio margins, we can derive a closed-form solution, as the margin requirement is the option's value plus a constant amount. For call options under the minimum portfolio margins, we derive the following upper bounds.

Corollary 5 *At time t, the upper price bound for a European call option with maturity T, strike K, and subject to CBOE's minimum portfolio margining rule is*

$$\begin{aligned} V_{call}(t) &= S(t)e^{(r_b - r_d - r_l)(T-t)}N(d_1(T,t)) - Ke^{-r_l(T-t)}N(d_2(T,t)) \\ &\quad + \frac{37.5(r_b - r_l)\left(1 - e^{-r_l(T-t)}\right)}{r_l} \end{aligned}$$

with

$$d_{1,2}(u,t) = \frac{\ln(\frac{S(t)}{K}) + (r_b - r_d \pm \frac{1}{2}\sigma^2)(u-t)}{\sigma\sqrt{u-t}},$$

where $N(\cdot)$ denotes the standard normal cumulative distribution function.

Analogously, we obtain the closed-form upper bound price for put options under the minimum portfolio margin requirement.

Corollary 6 *At time t, the upper price bound for a European put option with maturity T, strike K, and subject to CBOE's minimum portfolio margining rule is*

$$\begin{aligned} V_{put}(t) &= Ke^{-r_l(T-t)}N(-d_2(T,t)) - S(t)e^{-r_d(T-t)}N(-d_1(T,t)) \\ &\quad + \frac{37.5(r_b - r_l)\left(1 - e^{-r_l(T-t)}\right)}{r_l} \end{aligned}$$

with

$$d_{1,2}(u,t) = \frac{\ln(\frac{S(t)}{K}) + (r_l - r_d \pm \frac{1}{2}\sigma^2)(u-t)}{\sigma\sqrt{u-t}}.$$

2.4 Numerical Illustration

To investigate the magnitude of the impact of funding costs and margin rules on option pricing bounds, we compute the option prices for realistic parameter values. Using the sample ranging from January 2002 to August 2010, we compute the average three-month Overnight Index Swap rate (OIS rate) and use this value as proxy for the lending rate ($r_l = 2.3\%$). The average three-month US-dollar Libor rate is used as the borrowing rate ($r_b = 2.6\%$). For the volatility parameter, we take $\sigma = 15\%$. We note that these parameter values are not representative for the period of an ongoing crisis. They may hold under normal market conditions. Furthermore, we use the Libor rate as the proxy for the borrowing rate. Hence, the spread we impose for our numerical analysis is a conservative estimate. Finally, we impose the margin parameters set by the CBOE for index options, i.e., we use $a_1 = 0.15$ and $a_2 = 0.1$ for the strategy-based margin. For the portfolio margin rules, the simulated market moves are 15 possible moves ranging from -8% to 6%.

2.4.1 The Impact of Margin Requirements on Option Price Upper Bounds

To measure the impact of margin requirements on option price upper bounds, we first plot in Figure 2.2 Panel A and B the percentage difference between the upper price bound under the CBOE margin rules and the Black–Scholes price for puts and calls with a three-month maturity. As input for the classical Black-Scholes model we use the lending rate as interest rate. The resulting Black-Scholes option price serves us as benchmark. The presence of funding costs and margin requirements causes a sizable increase in the option prices over those implied by the Black–Scholes model. The relative price difference is convex and increasing in the strike price for call options and decreasing in put options. This effect is most pronounced for OTM options. Among the three margin rules we consider, the price increase is largest for strategy-based margins for a naked short sale, echoing the fact that they are the most stringent margin rules. For call options with moneyness $K/S = 1.2$, the price given by our model is 32% higher than the Black–Scholes price. For put options with moneyness $K/S = 0.8$, the relative increase due to margin requirements amounts to roughly 270%. For portfolio margining rules, the two types of margin rules generate very similar price increases, the magnitude of which is much smaller than the price increase we observe for the strategy-based margin for a naked short sale. However, for OTM calls the difference is still around 5 percent for moneyness $K/S = 1.2$ and 13 percent for put options with moneyness $K/S = 0.8$. Therefore, even under normal market conditions, funding costs and margin requirements have a non-negligible effect on option pricing bounds.

The relatively large impact of margin requirements on OTM options in Figure 2.2 arises because, in absolute terms, the collateral requirement could be substantial for deep OTM options, which have only small market value. For example, the margin specified by the CBOE for calls under the strategy-based margin rules satisfies $C(t) \geq a_2 S(t) + V(t)$. Under the portfolio

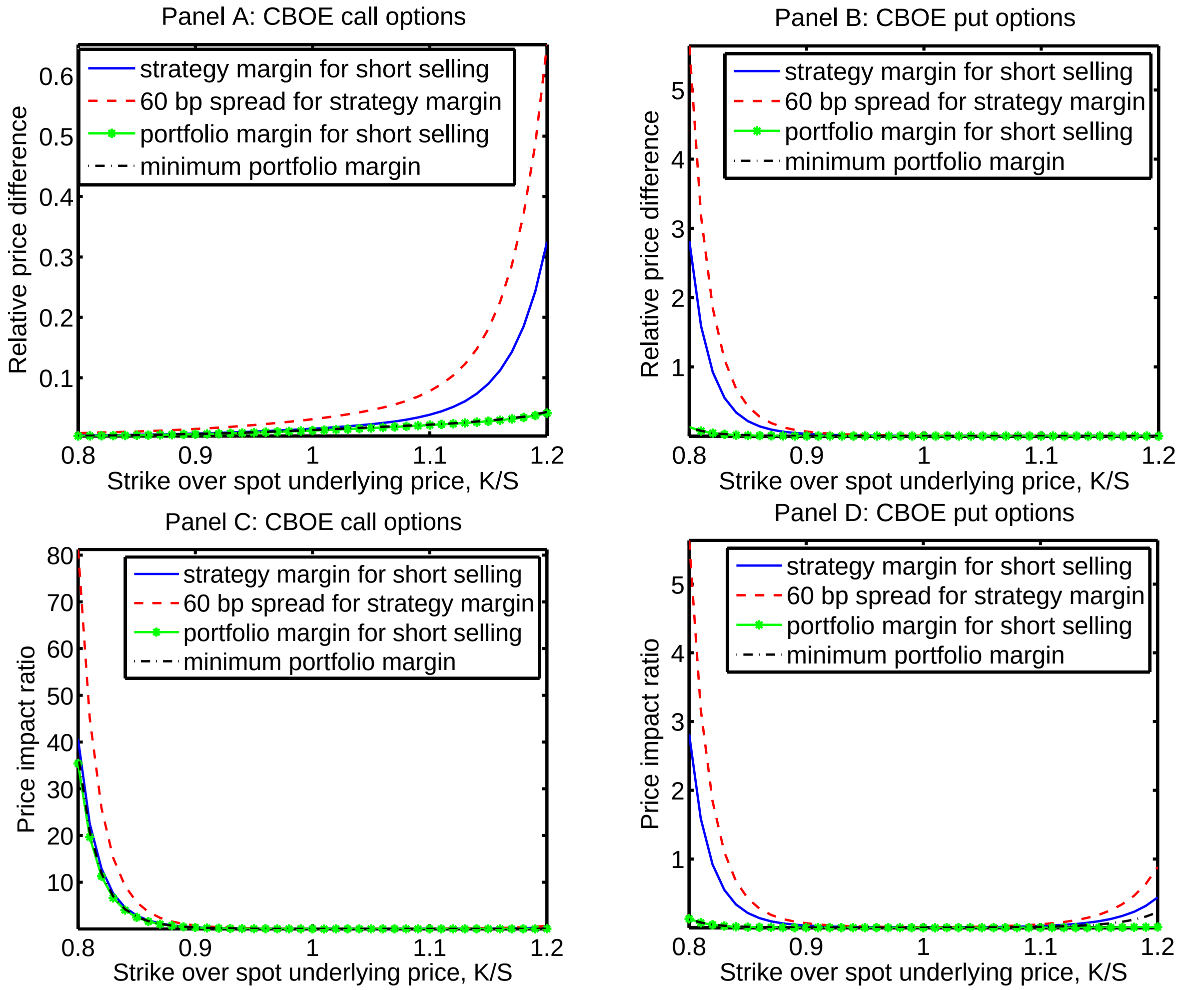

FIGURE 2.2: *Impact of margin requirements and funding costs on option price upper bounds.* We plot the percentage price differences between the Black–Scholes model and upper bounds derived from the margin model for call options (Panel A) and put options (Panel B) traded on the CBOE. Panels C and D show the price impact ratio defined in equation (2.7) for call options and put options, respectively. The parameter values with 30 bps funding cost are $r_b = 0.026$, $r_l = 0.023$, $\sigma = 0.15$, $a_1 = 0.15$, $a_2 = 0.1$, $T = 0.25$. To generate 60 bps funding cost, we hold the lending rate constant and increase the borrowing rate to $r_b = 0.029$. Furthermore, the underlying index level is assumed to be 1000. Per contract minimum margin 37.50 is applied for options with a multiplier of 100.

margining rule, $C(t) \geq 37.50 + V(t)$. Therefore, the size of the collateral relative to the option price may become substantial for small option values.

To give a more symmetric depiction of the impact of margin requirements on the component of an option's value that is determined by volatility, we remove the option's intrinsic value from our analysis and define the following quantity, which we call the price impact ratio:

$$\text{Price impact ratio} = \frac{\text{Option price upper bound} - \text{Black-Scholes Price}}{\text{Black-Scholes Price} - \text{Option's intrinsic value}}, \tag{2.7}$$

where the intrinsic value is defined as $\max\{0, S - Ke^{-rt}\}$ for calls and $\max\{0, Ke^{-rt} - S\}$ for puts. In Panels C and D of Figure 2.2 we plot the price impact ratio for calls and puts. The price impact ratio is a decreasing function of strikes for call options and a convex function for put options. Margin requirements have the highest impact on options with low strikes. For call options with moneyness $K/S = 0.8$, accounting for margin requirements generates a price impact ratio between 35 and 40, depending on the type of margin requirement. For put options with moneyness $K/S = 0.8$, the price impact ratio increases to 2.8 for the strategy margin, while for portfolio margin requirements the ratio increases to 0.12 only.

To examine the sensitivity to funding costs, Figure 2.2 also plots the resulting option price upper bounds when funding costs rise to 60 bps for the strategy-based margin for a naked short sale. An increase in funding costs leads to a larger price increase across all levels of moneyness. The assumption of a spread as large as 60 bps might seem excessive. However, we recall that during the recent crisis, the Libor–OIS spread peaked significantly over 300 bps and averaged nearly 100 bps between August 2007 and March 2009.

2.4.2 Margin Requirements and Implied Volatilities

So far, our results have demonstrated that funding costs and margin requirements have a sizable impact on the upper bounds for option prices. We now investigate the potential impact of these market frictions on the volatility smile. Since there is a one-to-one correspondence between IV and option prices, the no-arbitrage band derived in Section 2.3 implies a no-arbitrage region for implied volatilities. Our aim is to find out whether market frictions such as funding costs and margin rules provide room for rationalizing volatility smile patterns documented in the literature, even under the assumption of constant volatility.

Panel A, C, and E in Figure 2.3 show the call options' IV bounds when the three different CBOE margin rules are taken into account. We choose options with maturities of one, three, and six months. The lower IV bound degenerates to a constant, as it is given by the standard Black–Scholes IV using the lending rate, our benchmark Black-Scholes price.

The upper IV bound for calls is a decreasing function of the strike price for ITM options. For short-dated options, implied volatility starts to increase again when the option turns OTM. Hence, the IV bound for calls exhibits skew and smile patterns as observed in the market.

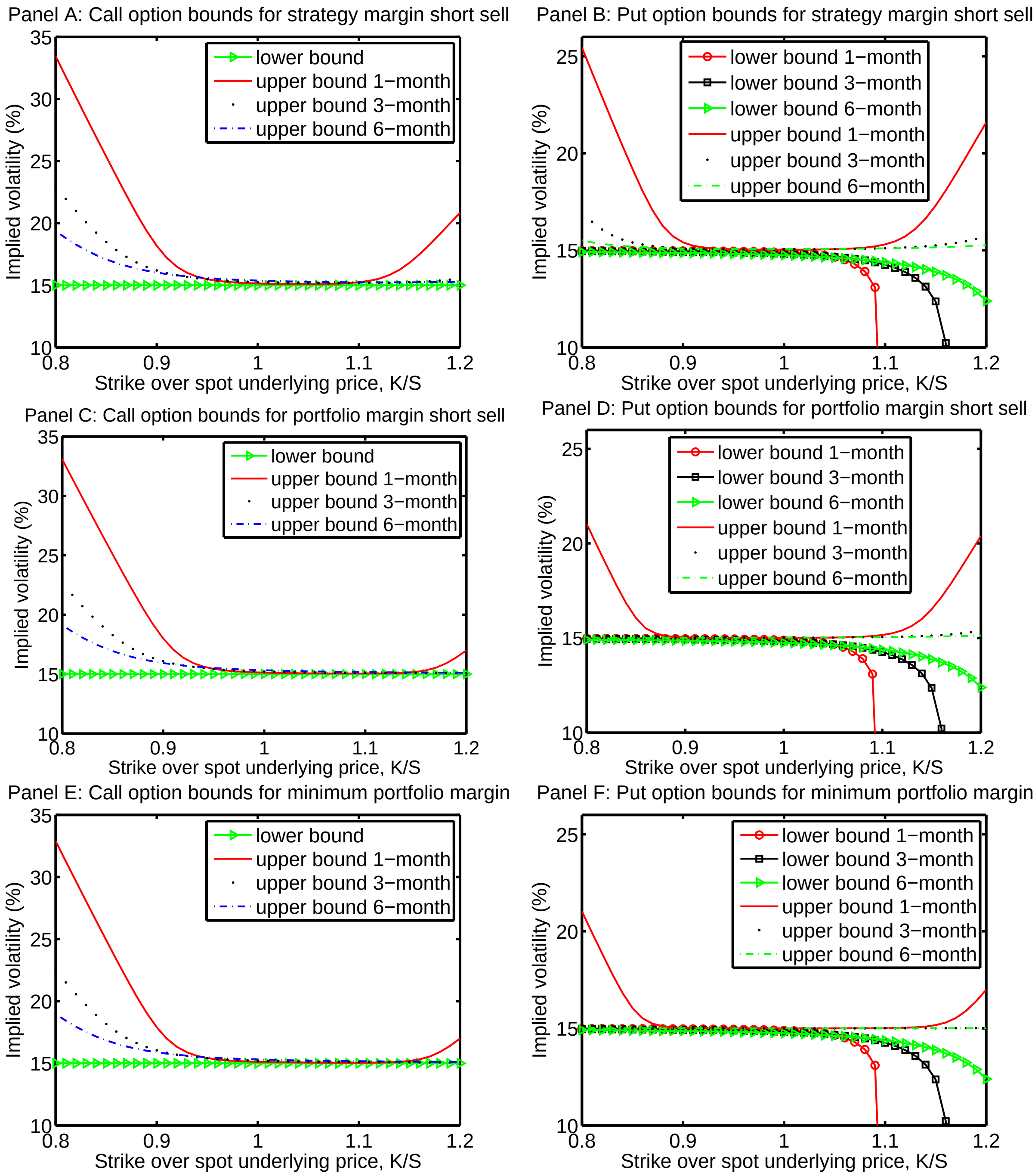

FIGURE 2.3: *No-arbitrage bounds for implied volatility (IV) curves.*
We plot the upper and lower bounds of the IV for options traded on the CBOE with one-month, three-month and six-month maturities for strategy-based margin requirements for a naked short sale, the portfolio margining requirements for a naked short sale and minimum portfolio margining requirements. Parameters $r_b = 0.026$, $r_l = 0.023$, $\sigma = 0.15$, $a_1 = 0.15$, $a_2 = 0.1$. We plot IV curves for calls in the left column and puts in the right column. For call options, the lower bounds collapse to a constant, i.e., to $\sigma = 0.15$, for all maturities and levels of moneyness.

Furthermore, consistent with previous empirical findings, IV curves generated by the model are steepest for one-month options, and gradually flatten out as maturity increases. Comparing the IV curves for the three margin rules, we find only small difference. The reason is that for call options, the price increase due to the replicating strategy involving buying is much more pronounced unless call options go deeply OTM. Therefore, the IV curves exhibit similar skew for three types of margin in our study.

For put options, the IV bounds exhibit a different pattern. In Panel B, D, and F of Figure 2.3, we plot the IV region for put options. The lower bound for puts is the Black–Scholes price using the borrowing rate. Hence, the lower bound is below the classical Black–Scholes price when the lending rate is used. Therefore, we obtain a downward sloping lower bound for IV, which becomes smaller than the value we fixed for the Black–Scholes volatility ($\sigma = 15\%$). For the upper bound, we also observe a volatility smile, which gradually flattens out as the maturity increases. Furthermore, the effects seem to be more sensitive to the margin rules applied. The slope accounting for the strategy-based margin is the steepest, while the two types of portfolio margin rules generate similar smile patterns.

The observed IV shape for call and put options is consistent with Panels C and D of Figure 2.2. Low strike options have higher price impact ratios. The price impact ratio measures the fraction of the upper bound price increase from the Black-Scholes price compared to the option's time value. Since the time value of options is largely affected by volatility, a higher price impact ratio is associated with a larger change in IV.

In Figure 2.3, we observe that the impact of funding costs on the upper bound of the IV surface is less pronounced for puts than for calls for the three types of margin requirements considered. This property is induced by the lower impact of funding costs on put options. Compared to a call, less borrowing is involved in replicating a put. Even though for both types of options the amount $V(t) - C(t)$ is borrowed from the debt account, the strategies on the underlying are different. For calls, investors borrow to purchase the underlying. In contrast, for put options, investors actually profit from selling the underlying short. Therefore, funding costs increase the replicating cost of calls to a greater extent than for put options.

We recall that the plots in Figure 2.3 represent upper and lower bounds. Hence, our results do not suggest that collateral requirements will indeed lead to higher IV for call options than for put options. We would have to add more structure to the model to provide sharper bounds. However, such extensions are beyond the scope of the present paper. Nevertheless, it is worth mentioning that Bollen and Whaley (2004) find that there is more demand for OTM index put options than OTM index call options. Consequently, market makers are likely to have larger net short positions in OTM put options. As portfolio margin rule uses as collateral the highest possible loss for the whole portfolio, selling put options is likely to put more collateral constraints on market makers. In contrast, selling call options might incur only the minimum collateral requirement. Therefore, the impact of collateral on OTM put options could be higher than on call options leading to higher IV of put options. Moreover, as put-call parity

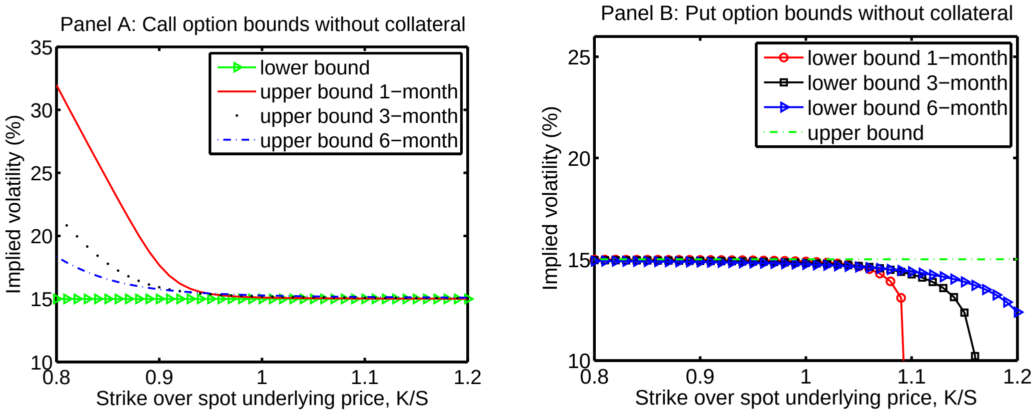

FIGURE 2.4: *No-arbitrage bounds for implied volatility (IV) curves.*
We plot the upper and lower bounds of the IV for options without margin requirements as predicted by Bergman (1995). Parameters $r_b = 0.026$, $r_l = 0.023$, $\sigma = 0.15$. We plot IV curves for calls in the left column and puts in the right column. In case of call options, the lower bounds collapse to a constant, i.e., to $\sigma = 0.15$, for all maturities and levels of moneyness. For put options, the upper bounds are constant at $\sigma = 0.15$.

does not necessarily hold in the empirical data (see Kamara and Miller (1995) for example), we do not expect call and put options with the same strike to have exactly the same IV.[7]

For comparison, in Figure 2.4, we plot the IV curves when only funding costs but no margin requirements are taken into account as in Bergman (1995). Excluding margin requirements, we can still observe a smile for call options. But for put options, the upper bound is exactly the benchmark Black-Scholes price. Thus, the upper bound degenerates to a constant. Hence, in a setting with constant volatility and funding costs, but without collateral requirements as in Bergman (1995), there is no way to explain the typical smile pattern observed for put options.

2.5 The Impact of Collateral under Stochastic Volatility

In the previous analysis, we have assumed constant volatility for the underlying price process. To allow for a richer dynamic process, we can analyze an obvious extension of our model and introduce stochastic volatility as in Heston (1993). Specifically, we assume that the asset price

[7] Applying put-call parity by recognizing the effect of margin requirements and funding costs gives another price bound. For example, given the model-implied call price bound, we can use put-call parity to obtain another put option price bound. The intersection of the put option price bound derived from the model and put-call parity gives a new bound. Numerical results show that this bound is wider than the options' own model-implied price bound. Therefore, put-call parity does not imply a sharper bound. The reason is that under funding costs and margin requirements, put-call parity generates two different inequalities.

follows:

$$
\begin{aligned}
dS(t) &= \mu S(t)dt + S(t)\sqrt{v(t)}dW^1(t), \\
dv(t) &= \kappa(\theta - v(t))dt + \xi\sqrt{v(t)}dW^2(t), \quad dW(t)^1 dW(t)^2 = \rho dt.
\end{aligned}
$$

Since we rely on a replication argument for deriving our option price bounds under collateral requirements, we need to introduce an additional volatility-sensitive asset, which we denote by $g(t)$. For simplicity, we assume that $g(t)$ is a variance swap. Hence, the replicating trading strategy is a five-dimensional process, which we denote by $(\alpha(t), \beta(t), \lambda(t), \delta(t), \gamma(t))$ to capture the holding of the stock, the cash borrowed, the cash deposited, the collateral, and the holding of the variance swap.

The collateral imposed on the hedging strategy depends on the riskiness of the whole strategy, and thus also on the variance swap. Since we do not know the exact form of the collateral requirement of the trading strategy, we can only make an educated guess about the actual size of collateral requirements. The fact that option buyers do not need to post collateral suggests that the lower bounds of option prices are not affected by the margin requirement of options. However, trading the variance swap might incur collateral, which has an impact on the option's lower bounds. For the sake of simplicity, for the lower bound we do not consider the collateral of the variance swap. Thus the collateral for the whole strategy is set at zero for lower bounds. Since posting collateral is always costly in the presence of funding costs, the collateral on the variance swap will decrease the lower bounds.

For the upper bound, selling options incur margin requirements. Santa-Clara and Saretto (2009) estimate the margin imposed on selling an option is usually more than 3.6 times and, for OTM options, up to 108 times the option value. Therefore, in order to have a conservative estimate of the impact of collateral and funding costs, we assume in our numerical example that the collateral for the whole trading strategy is twice the value of option prices.[8]

The option price bounds are derived by solving similar $\mathscr{M}^-$ and $\mathscr{M}^+$ problems as in Section 2.3, but with more involved spot price dynamics and an additional hedging instrument in the replicating portfolio to hedge against changes in volatility. Solving similar PDEs as in the standard Heston model, we obtain the option price bounds in the presence of collateral requirements. We present the derivation of the PDE in Appendix 2.C. To solve the model numerically, we use finite difference methods.

Figure 2.5 plots the upper and lower bounds for a given set of parameters. The IV upper bounds for calls (Panel A) and puts (Panel B) suggest that collateral and interest spreads could drive up the IV significantly. The lower bound of call options corresponds exactly to the standard Heston model (Heston IV). However, for put options the Heston model does not

[8] Santa-Clara and Saretto (2009)'s estimates are based on strategy margin. But even in the presence of portfolio margin, the margin would exceed the value of the option. Therefore, $C(t) = 2V(t)$ is still a conservative estimate.

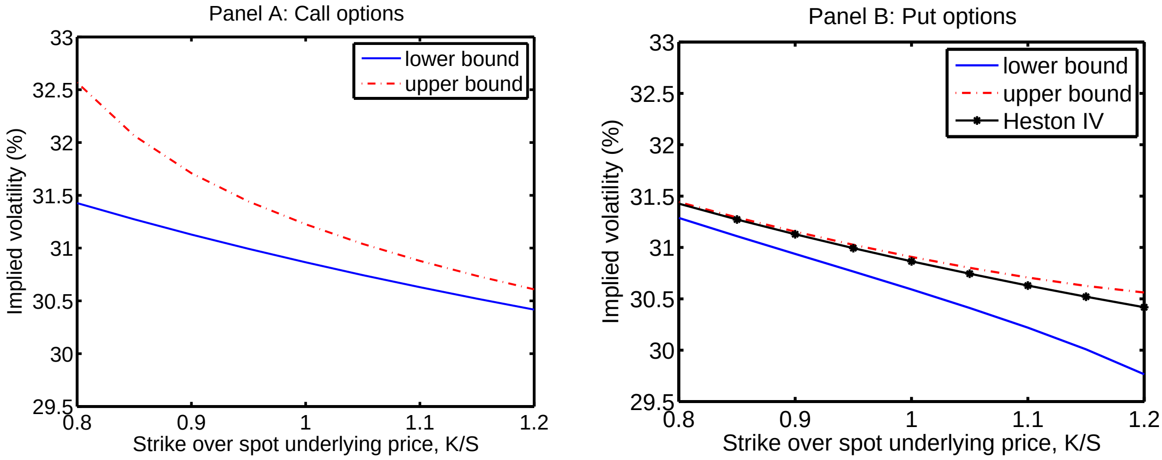

FIGURE 2.5: *IV no-arbitrage bounds when the underlying has stochastic volatility*
We plot the upper and lower bounds of the IV for options with six-month maturities. We fix the parameters as $r_l = 0.023$, $r_b = 0.026$, $\xi = 0.1$, $\theta = 0.01$, $\kappa = 3$, $\rho = -0.5$, $v_0 = 0.25$, $C(t) = 2V(t)$, and $m^* = 2$. We plot IV curves for calls in the left column and puts in the right column. For calls, the lower bounds are the same as for the standard Heston model (Heston IV).

coincide with the lower bound. The prices of the Heston model lie in the region spanned by the lower and upper bound. Moreover, we observe a slightly higher IV upper bound for call options than put options. This observation is consistent with the case of constant volatility and is due to the fact there is more borrowing involved in replicating a call option.

The asymmetry between margin requirements for long and short positions has a large impact on the option bid-ask spreads. Market makers post little margin when buying an option from someone who sells at the bid. However, they have to post substantial margin when selling an option to someone who buys at the ask. To some extent, we can interpret the upper and lower bounds as option bid-ask spreads implied by our model. Under the assumption of stochastic volatility, our model shows a similar IV shape observed in the market that both bid and ask price yield a smile, with a stronger smile for ask prices.[9]

Since margin requirements have no impact on the lower bound, we analyze the sensitivities of upper bounds to changes in collateral and interest spreads.[10] In Figure 2.6, we report the IV upper bounds when the borrowing rate r_b varies (Panels A and B) and when the collateral $C(t)$ varies (Panels C and D). As expected, IV upper bounds increase with the borrowing rate and collateral requirements. When the spread of interest rates is as high as 3 percent, a typical value during the recent financial crisis, collateral requirements could increase the IV

[9] In contrast to the constant volatility case, both upper and lower bounds in the stochastic volatility model exhibit a typical IV smile pattern. However, our model does not necessarily imply that bid-ask spread across moneyness must have a shape as in Figure 2.5, since we used a simple uniform haircut for all levels of moneyness.

[10] Option price bounds are not very sensitive to changes in the other parameters. Hence, we do not plot the changes of the upper bounds due to changes in the other parameters.

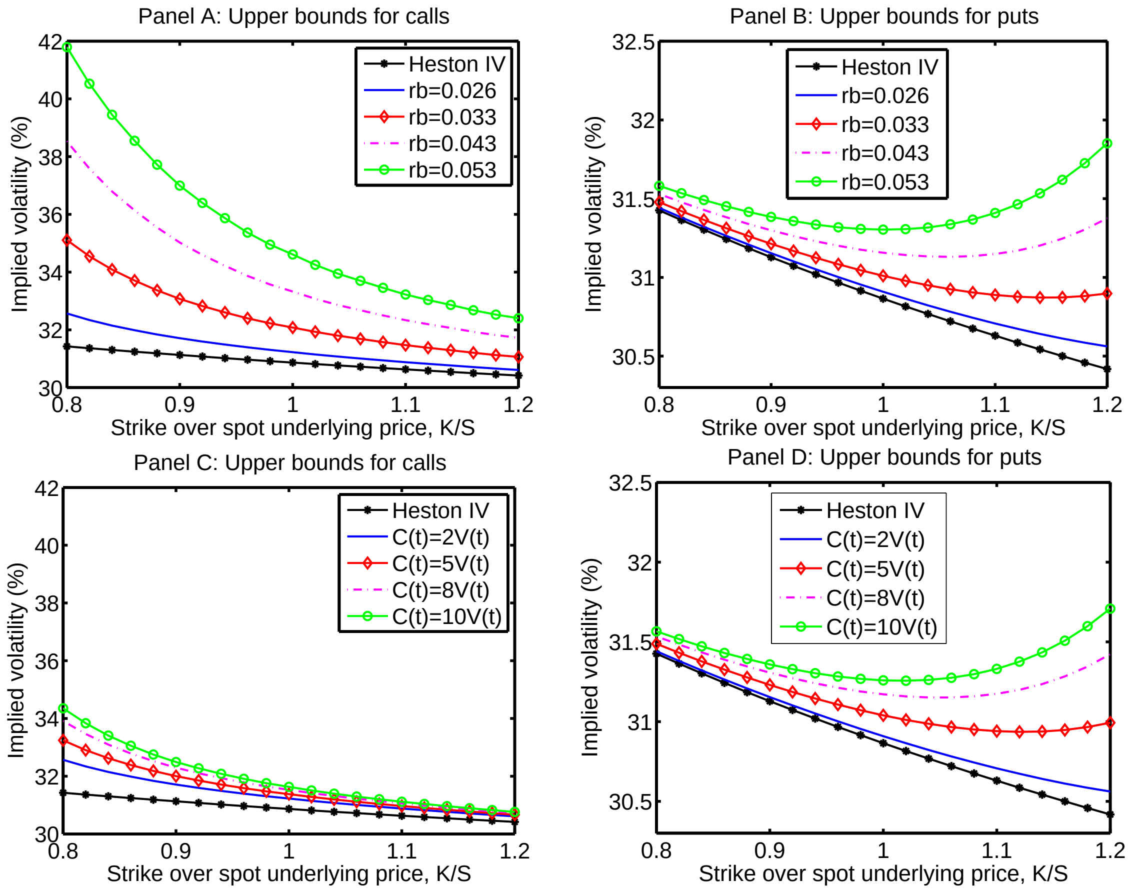

FIGURE 2.6: *IV upper bounds under different borrowing rates and margin requirements*

In Panels A and B, we plot the upper bounds of the IV for options with six-month maturities when the borrowing rate changes. In Panel C and D, we plot the upper bounds of the IV for options with six-month maturities when the margin requirement changes. For the parameters, we use $r_l = 0.023$, $r_b = 0.026$, $\xi = 0.1$, $\theta = 0.01$, $\kappa = 3$, $\rho = -0.5$, $v_0 = 0.25$, $C(t) = 2V(t)$, and $m^* = 2$. We plot IV curves for calls in Panels A and C, and puts in Panels B and D.

substantially. For ITM call options with moneyness 0.8, the IV can jump from around 31.5% for the standard Heston model to around 42%. For put options, the increase is less dramatic. For a moneyness of 1.2, the IV jumps from under 30.5% to around 32%.

If we compare the volatility increase across different levels of moneyness, we see that for put options the presence of collaterals may even lead to a veritable volatility smile in that the upper bounds are higher for ITM puts than for ATM puts. When we assume an extreme wedge between borrowing and lending rate, the ITM IVs may be even higher than the OTM IVs. This feature follows from our assumption about the collateral. We assume a constant margin haircut across moneyness measured as the margin requirement divided by the value of the option. Santa-Clara and Saretto (2009) find that OTM options have a much higher margin hair cut, i.e., the ratio of the margin requirement to the value of the option is much higher for OTM put options. If we were to impose this more realistic margin requirement, we would see again an IV upper bound for put options that decreases with increasing moneyness.

We remark that the bounds derived for the stochastic volatility setting are only approximative. Given the parameter choice, our estimators are conservative in that they underestimate the effect of collateral requirements. As we require an additional instrument in building up our replicating strategy, we have no idea about the sign of the cash flow. We can only approximate the upper bound by applying the borrowing rate to the cash flows of the hedging portfolio. Therefore, our model's upper bound must be below the true upper bound.

In contrast to the stochastic volatility model, the assumption of constant volatility allows us to fully account for the impact of collateral and funding costs, since we can apply the proper interest rates to every trade in the replication strategy. Given these obstacles of the stochastic volatility model, we decide to focus on the constant volatility case for our empirical application. By doing so, we also circumvent the problem of estimating the dynamics of the variance swap and the specification of the variance swaps' collateral requirement. We leave these challenging questions for future research.

2.6 Empirical Application

Our intention was to build a simple model to isolate the effect of funding costs and collateral requirements. As the volatility surface could be characterized by IV slopes and levels, we could challenge our model by comparing these quantities with the ones implied by the upper price bounds. Since it is clear from the analysis in the previous section that the effect of funding costs shows up prominently across the moneyness dimension, we will only conduct an empirical analysis on IV slopes but not IV levels. If our model could add additional explanatory power to the factors previously used in the finance literature for describing the IV slope, our finding would provide a strong argument to incorporate funding costs and differential interest rates in option pricing models.

2.6.1 Data Processing

We use data for options written on SPX from Ivy DB OptionMetrics for the sample period ranging from January 2002 to August 2010.[11] Our data covers the period of the financial crisis. The high funding costs during this period provide ideal data points for our test. End-of-day bid and ask quotes, open-interest, volume, exercise price, IV, delta, gamma, dividend yield, and expiration dates on every call and put option are all provided by OptionMetrics.

Several filtering rules have been applied to obtain a clean data set. Firstly, we eliminated options with maturities less than eight days or more than 150 days to exclude any liquidity-related bias. Secondly, we included only options with a positive trading volume, positive open interest, and positive bid prices. Finally, mid quotes lower than 0.375, bid–ask spreads more than 1.5 times the mid-quotes, and strike over spot prices less than 0.7 or more than 1.3 were also excluded. This data contains in total 153,926 calls and 198,775 puts.

The lending rate is proxied by the US OIS rate and the US interbank borrowing rate is captured by the Libor rate. The interest rates are obtained from Bloomberg.[12] To obtain the interest rates at different maturities we use linear interpolation.

As an alternative, we could also use the put-call parity (PCP) to back out the borrowing and lending rate from options on SPX as was done, e.g., in Brenner and Galai (1996), Jackwerth and Rubinstein (1996), and Constantinides, Jackwerth, and Perrakis (2009), among others. However, in untabulated results, we find that our analysis and findings do not change and are robust to different specifications of the interest rates. As a further model input, we require a proxy for volatility, for which we use the VIX.[13]

As our model predicts different shapes for the put and call IV curve, we estimate the empirical slopes for puts and calls separately. When computing the IV, the midpoint of the best closing bid price and best closing offer price for the option is used. Following the standard methodology of BKM (2003), we derive the slope estimates weekly by pooling all the IV data in any given week (449 weeks) from Wednesday to Tuesday. We then sort the options according to their time-to-maturity into two groups, short-term options (maturity of less than 60 days) and medium-term options (maturity between 60 and 150 days).

In addition to analyzing the whole sample period from January 2002 to August 2010, we also perform our tests on two subperiods including the pre-crisis period January 2002 to July 2007 and the crisis period August 2007 to July 2009. The results for these subperiods are similar to

[11]We focus on options on the SPX, as they are European options. Exchange-traded single stock options are of the American type and, hence, would complicate our analysis.

[12]Note that even though OptionMetrics provides index put options prices traded from 1996, the OIS rate from Bloomberg is only available since the end of 2001. Therefore, we select the data sample from 2002. Furthermore, using other proxies for the lending rate such as US Treasury rates produces similar results. The US Government began issuing four-week Treasury bills in mid 2001. Therefore, using the Treasury rate as the lending rate would not significantly extend our sample period.

[13]Initially, we considered three distinct volatility measures: 30-day historical volatility, VIX, and the IV of the closest ATM 1-month options. The key results are robust and remain unchanged for these different measures.

the whole sample. Therefore, they are omitted for brevity but can be obtained by the authors.

2.6.2 Regression Results for IV Slopes

To derive the slope estimate, different measures for the slope of the IV curve have been suggested in the literature. Here, we follow BKM (2003). They estimate the slope coefficient Π using the regression equation

$$\ln(\sigma^{iv}(y_j)) = \Pi_0 + \Pi \ln(y_j) + \varepsilon_j, \quad j = 1, ..., J, \tag{2.8}$$

where y denotes the moneyness K/S, σ^{iv} denotes the Black–Scholes IV and J is the number of options available in the week. We perform the regression in (2.8) for each maturity group of call and put options to obtain weekly slope estimates.

Table 2.1 reports the estimated slope coefficients and the corresponding R^2 for each option category. The slopes are negative and more pronounced for short-term options with slightly more negative slopes for calls than for puts. For each option group, the regression in equation (2.8) captures between 70-90% of the variation in the IV slope. The slope for call options seems to be more negative than for put options, a result which is due to log transformation in regression (2.8). Indeed, not taking the logarithm in (2.8) or using other definition of slopes such as, e.g., in Han (2008), gives more negative slopes for puts.[14] The empirical slopes we obtained show a strong persistence over time. Running augmented Dickey-Fuller test and choosing the number of lags by the Akaike information criterion suggests that slopes for all categories are non-stationary I(1) processes.

[Table 2.1 about here]

To obtain our model-implied slope Π^{model} under constant volatility, we first compute the upper bounds of put and call options at nine different equally-spaced moneyness levels K/S ranging from 0.8 to 1.2. We convert these prices to Black-Scholes IVs, which we then use for running the regression in equation (2.8). We use three types of margin rules to obtain the option upper bound and derive model-implied slopes. The following analysis is conducted for slopes derived using three margin rules.

To avoid the problem of spurious regression, we take the differences of all variables for the regression. Firstly, we run the $\Delta\Pi_t$ on its lag $\Delta\Pi_{t-1}$ as follows

$$\Delta\Pi_t = \beta_0 + \beta_1 \Delta\Pi_{t-1} + \varepsilon_t. \tag{2.9}$$

To see whether our model-implied slopes could explain the time variation of the empirical

[14] In unreported regressions, we also used the slope definition from Han (2008), where the slope is measured as the negative of the average OTM put IV over the average ATM put and call options IV. The results are similar to what we find using the BKM (2003) slope definition.

slopes, we regress the empirical slope change on the model-implied slope change

$$\Delta\Pi_t = \beta_0 + \beta_1 \Delta\Pi_t^{\text{model}} + \beta_2 \Delta\Pi_{t-1} + \varepsilon_t, \tag{2.10}$$

where the lagged slope difference $\Delta\Pi_{t-1}$ is included in the regression to correct for the autocorrelation in the dependent variable. We present the results from regressions (2.9) and (2.10) in Table 2.2 for different option groups.

[Table 2.2 about here]

Table 2.2 gives us several interesting findings. The lagged empirical slope difference, although always significant, can only explain a small portion of the evolution of $\Delta\Pi_t$ with average R^2 around 5 percent. For regression (2.10), we find that the coefficient for $\Delta\Pi_t^{\text{model}}$ for all margin rules is always positive and significant at the 1% level, indicating a positive link between the empirical slope difference and our model-implied slope difference. The coefficients for the put options are larger than those for the call options. This observation is in line with the findings of our numerical investigation in Section 2.4, where we find a steeper IV curve for calls as they are more sensitive to funding costs. The coefficient for calls does not differ much for different margin rules, which is again consistent with the finding in Section 2.4 that similar smiles are observed for call options.

For puts, however, we do observe quite different coefficients. As shown in our numerical analysis in Section 2.4, strategy-based margins tend to generate a steeper IV smile. Therefore, the coefficient is relatively small for strategy-based margins. Moreover, the coefficient for the minimum portfolio margins is also small compared to the portfolio margin for a naked short sale. The minimum portfolio margin tends to increase OTM IVs much more than ITM IVs, as the per contract minimum is substantial only for OTM options. In contrast, the naked short sale portfolio margin rules raise the IV of options across all moneyness levels, giving rise to a flatter smile. Therefore, the coefficient for the naked short sale portfolio margining is much higher than for the other two margin rules.

Finally, we see that for both puts and calls our model-implied slope can generate adjusted R^2-values around 23.6 percent for short-term options and around 39.1 percent for medium-term options. These findings provide evidence that our model helps to explain a substantial part of the time variations of empirical IV slope differences.

2.6.3 Regression Results Including Control Variables

To compare the performance of regression (2.10) with those of other models, we also provide a regression analysis including other control variables. As a first set of control variables we consider the risk-neutral skewness and kurtosis. As shown by BKM (2003), the second and third moments of risk neutral distribution of returns have significant explanatory power in describing the time variation of empirical slopes.

As a second set of control variables, we consider the following three commonly used variables. We include the VIX as a proxy for market volatility, the previous six-month returns to capture stock market momentum, and a relative demand factor to control for demand impact.[15] In addition, since our model implies that funding costs matter for the slope of IV curves, we also include Libor–OIS spreads in our regression.

We start with the following specification of the regression equation based on risk-neutral parameters:

$$\Delta\Pi_t = \beta_0 + \beta_1 \Delta\text{Skewness}_t + \beta_2 \Delta\text{Kurtosis}_t + \beta_3 \Delta\Pi_{t-1} + \varepsilon_t. \tag{2.11}$$

As an additional exercise, we combine our model-implied slopes with risk-neutral parameters in one single regression as follows:

$$\Delta\Pi_t = \beta_0 + \beta_1 \Delta\text{Skewness}_t + \beta_2 \Delta\text{Kurtosis}_t + \beta_3 \Delta\Pi_t^{\text{model}} + \beta_4 \Delta\Pi_{t-1} + \varepsilon_t. \tag{2.12}$$

We run this regression for all of the three types of margin rules discussed in Section 2.2. We report the results for regressions (2.11) and (2.12) in Table 2.3.

[Table 2.3 about here]

For regression (2.11), we observe in Table 2.3 that the risk-neutral skewness is not significant at the 5 percent level for any option group. The risk neutral kurtosis becomes significant for the medium-term option group only. The lagged slope difference is always significant at any reasonable statistical level. However, using risk neutral factors alone gives quite low R^2 values. In the combined regression (2.12), we observe that the model-implied slope differences are significant at the 1 percent level for all types of margin rules and all option groups. The risk neutral factors remain insignificant for all option groups. The adjusted R^2 values have improved considerably by adding model-implied slope differences.

For the second set of control variables, we first run the following regression with control variables only,

$$\begin{aligned}\Delta\Pi_t \;=\; & \beta_0 + \beta_1 \Delta\text{LiborOIS}_t + \beta_2 \Delta\text{VIX}_t + \beta_3 \Delta\text{IndexReturn}_t \\ & + \beta_4 \Delta\text{RelativeDemand}_t + \beta_5 \Delta\Pi_{t-1} + \varepsilon_t. \end{aligned} \tag{2.13}$$

[15]These variables are used to explain the time variations of the slope of the IV curves by, e.g., Amin, Coval, and Seyhun (2004), Li and Pearson (2005), Bollen and Whaley (2004) and Garleanu, Pedersen, and Poteshman (2009). Unfortunately, we do not have access to the data to measure the demand impact of end users as in Garleanu, Pedersen, and Poteshman (2009). We follow Han (2008) to measure the demand impact by the ratio of total open interest for OTM index put options (defined by $-\frac{3}{8} < \Delta_P \leq -\frac{1}{8}$ where Δ_P is the delta of put options) to that for near and ATM index options (defined as call options with $\frac{3}{8} < \Delta_C \leq \frac{5}{8}$ and put options with $-\frac{1}{8} < \Delta_P \leq -\frac{3}{8}$ where Δ_C denotes the delta of call options).

Analogously, we also run a combined regression as follows,

$$\begin{aligned}\Delta\Pi_t \quad = \quad & \beta_0 + \beta_1 \Delta\text{LiborOIS}_t + \beta_2 \Delta\text{VIX}_t + \beta_3 \Delta\text{IndexReturn}_t \\ & + \beta_4 \Delta\text{RelativeDemand}_t + \beta_5 \Delta\Pi_t^{\text{model}} + \beta_6 \Delta\Pi_{t-1} + \varepsilon_t. \end{aligned} \tag{2.14}$$

We report the results for regression (2.13) and regression (2.14) in Table 2.4 for different options groups.

[Table 2.4 about here]

Referring to Table 2.4, we find that in regression (2.13), ΔVIX_t is significant at the 5 percent level for all option groups. The demand factor is only significant for short-term calls. All other control variables are not significant at the 5 percent level. In the combined regression (2.14), $\Delta\text{LiborOIS}_t$ and ΔVIX_t are not always significant. Their coefficients switch signs for different option groups. However, the significance of $\Delta\Pi_t^{\text{model}}$ remains at the 1 percent level, even after controlling for other variables. $\Delta\Pi_t^{\text{model}}$ changes from one week to the next because LiborOIS_t and VIX change. As a non-linear function of LiborOIS_t and VIX, changes in Π_t^{model} have additional power beyond that provided directly by changes in LiborOIS_t and VIX_t. Indeed, when we include the model-implied slopes, we can substantially increase the explanatory power. The residual effect of our model-implied slope after controlling for VIX and the Libor–OIS spread is positive, indicating that a higher implied slope change is followed by a higher empirical slope change. We remark that the above results are invariant to different margin requirements and hold for all option groups.

2.7 Conclusion

We presented a tractable option pricing model that accounts for margin requirements on exchanges and market participants' funding costs. In a dynamically incomplete market with differential rates, we derived upper and lower bounds for option prices with margin requirements when the underlying follows a geometric Brownian motion. Since margin requirements are positive, the prices derived from the upper bounds exceed the classical Black–Scholes option prices. For the margin rules of the world's most important option exchange, the CBOE, we derived upper price bounds for European call and put options. The relative difference between these upper bounds and the original Black–Scholes option prices turns out to be substantial, even under normal market conditions. Analyzing the funding costs in volatility space, the no-arbitrage region we obtained for the IV provides enough flexibility to allow volatility smiles and skews that are comparable in size to the empirically observed IV patterns. Consistent with empirical findings, the IV curve flattens out as the maturity increases. Hence, funding costs and collateral requirements offer an institutional explanation of the volatility smile phenomenon without departing from the constant volatility assumption. As an additional analysis, we also investigate the impact of margin requirements and funding costs when the underlying asset exhibits stochastic volatility. Our numerical analysis confirms that there is still a significant increase in the IV upper bound from the Heston IV due to margin requirements and funding costs.

The complexity of stock price processes and the variety of factors influencing option markets makes an empirical test of our model a delicate task. However, our model highlights that the slopes generated by the IV upper bounds under constant volatility assumption capture important factors in the time variation of the empirical slope change. By fitting the change of SPX option IV slopes, we found that our model-implied slopes are quite successful in explaining the empirical slopes, with average adjusted R^2 around 30 percent. The performance of our model-implied slope was compared with two regressions where risk-neutral factors and other commonly used variables are taken as the regressors. Using our model-implied slopes, we found that our institutional factors generate a level of adjusted R^2 much higher than the one generated by the commonly used factors. Furthermore, we ran a combined regression where both the model-implied slope and control variables are included. The regression results showed that our model-implied slopes remain significant and add significant explanatory power to the regression. Therefore, we conclude that our model, albeit simple, offers promising avenue for rationalizing the impact of margin requirements and funding costs on option prices.

Appendix

2.A Derivation of the Upper Price Bounds

We first derive the upper bounds for call options. We assume that the underlying price $S(t)$ follows a geometric Brownian motion with log-increments having constant volatility σ. Let $V(t)$ denote the upper bound of the derivative contract price. Applying Ito's lemma allows us to find the dynamics of $V(t)$:

$$dV(t) = \left(V_t'(t) + \frac{1}{2}\sigma^2 S^2(t) V_{ss}''(t)\right) dt + \alpha(t) dS(t),$$

where $\alpha(t) = V_s'(t)$. The option writer can construct a self-financing portfolio by holding $\alpha(t)$ units of stocks and taking positions in the debt, cash, and collateral accounts. We denote the corresponding amounts in these accounts by $\beta(t)$, $\lambda(t)$, and $\delta(t)$. Hence, the replicating strategy has a value $U(t) = \alpha(t)S(t) + \beta(t) + \lambda(t) + \delta(t)$, which should be equal to $V(t)$. As self-financing implies no injection of external capital, the dynamics of the hedging portfolio must be

$$dU(t) = \alpha(t)(dS(t) + r_d S(t) dt) + r_b \beta(t) dt + r_l \lambda(t) dt + r_l \delta(t) dt.$$

The total value of the accounts is the difference between the value of the strategy and the value of the purchased stocks, i.e., $\beta(t) + \lambda(t) + \delta(t) = V(t) - \alpha(t)S(t)$. In the classical Black–Scholes setting, this value would grow at the unique risk-free rate. However, in our model the lending rate determines the evolution of the cash and collateral account, while the borrowing rate determines the evolution of the debt account. Therefore, we must carefully segregate the positions into *i)* the collateral $C(t)$ required to be deposited in the cash account earning the lending rate, *ii)* the quantity $V(t) - C(t)$ borrowed at the borrowing rate from the debt account to finance the posting of the margin, and finally *iii)* $\alpha(t)S(t)$ borrowed from the debt account to finance the stock purchase.

Since the value of $C(t)$ is always greater than $V(t)$, the difference $V(t) - C(t)$ is negative and needs to be borrowed from the debt account. Summing up all positions in the debt and cash account and using the appropriate interest rates yields the following dynamics for the value of the accounts:

$$d\left(\beta(t) + \lambda(t) + \delta(t)\right) = (r_l C(t) - r_b (C(t) - V(t)) - r_b \alpha(t) S(t))\, dt.$$

Since the value of the replicating strategy equals the value of the derivative, the option value must satisfy the PDE

$$V_t'(t) + \frac{1}{2}\sigma^2 S^2(t) V_{ss}''(t) = r_b V(t) - (r_b - r_l) C(t) - (r_b - r_d)\alpha(t) S(t)$$

which we can rewrite as

$$V_t'(t) + (r_b - r_d)S(t)V_s'(t) + \frac{1}{2}\sigma^2 S(t)^2 V_{ss}''(t) = r_b V(t) - (r_b - r_l)C(t) \tag{2.15}$$

with the boundary condition

$$V(T) = (S(T) - K)^+. \tag{2.16}$$

The continuity of $C(t)$ allows us to make use of the Feynman–Kac Theorem to represent the solution to the PDE in (2.15) in terms of the following expectation:[16]

$$V(t) = \underbrace{\mathbb{E}_t^{\mathbb{P}_b}\left[e^{-r_b(T-t)}V(T)\right]}_{(A)} + \underbrace{\mathbb{E}_t^{\mathbb{P}_b}\left[\int_t^T e^{-r_b(u-t)}(r_b - r_l)C(u)\,du\right]}_{(B)}. \tag{2.17}$$

We note that the expectation in equation (2.17) is taken under that pricing measure $\mathbb{P}_b$ for which the stock price discounted by $r_b - r_d$ follows a martingale.

To replicate a put option, the investor has to short sell a certain amount of the underlying and invest it in the cash account. Hence we have three positions: *i)* the collateral $C(t)$ deposited in the cash account, *ii)* the quantity $V(t) - C(t)$ borrowed to finance the required margin, and *iii)* the short sell proceeds $\alpha(t)S(t)$ deposited in the cash account. For put options, the option's price is not sufficient to meet the margin requirement and $V(t) - C(t)$ needs to be funded by borrowing. The relative size of $\alpha(t)S(t)$ and $V(t) - C(t)$ is not known. Thus we assume that the short selling proceeds $\alpha(t)S(t)$ are saved in the cash account and could not be used to satisfy the margin requirement. This assumption not only simplifies the model, but is also consistent with market practice. Short sellers are generally required to leave the short sale proceeds in an interest bearing account with their broker until the short position is closed.[17] The total growth of the cash, debt, and collateral account is then equal to

$$d(\beta(t) + \lambda(t) + \delta(t)) = [r_l C(t) - r_b(C(t) - V(t)) - r_l \alpha(t)S(t)]\,dt.$$

Equating the replicating strategy value with the put option value $V(t)$ gives us the PDE for $V(t)$:

$$V_t'(t) + (r_l - r_d)S(t)V_s'(t) + \frac{1}{2}\sigma^2 S^2(t)V_{ss}''(t) = r_b V(t) - (r_b - r_l)C(t), \tag{2.18}$$

[16]We remark that the solution to equation (2.17) is indeed the solution to the $\mathcal{M}^+$ problem. It is the value of a self-financing strategy satisfying the collateral requirement of the option writers. Its payoff at time T is equal to the payoff of the call option. Furthermore, no simultaneous borrowing and lending in the debt and cash account is involved in the replicating strategy. Therefore, the initial investment cost is minimized.

[17] Bergman (1995) even discusses the case when brokers collect the interest rates to compensate for their own monitoring costs. In such a case, the replicating costs for put options are even higher. However, as we only consider two rates in our model, we keep the assumption that short selling earns the lending rate.

with the boundary condition

$$V(T) = (K - S(T))^+. \tag{2.19}$$

When $C(t)$ is continuous, we can alternatively represent the PDE by[18]

$$V(t) = \mathbb{E}_t^{\mathbb{P}_l}\left[e^{-r_b(T-t)}V(T) + \int_t^T e^{-r_b(u-t)}(r_b - r_l)C(u)du\right]. \tag{2.20}$$

We note that for put options, the underlying has a drift term $r_l - r_d$ under the pricing measure $\mathbb{P}_l$, compared with $r_b - r_d$ for calls. According to the Feynman–Kac formula, the drift term of the underlying under the risk-neutral measure is determined by the coefficient of $\frac{\partial V(t)}{\partial S}$ in the PDE. For puts, the short sale proceeds are invested at r_l while for calls, longing the underlying requires borrowing at r_b. Therefore, for puts and calls, different drift terms adjusting for dividends are applied to the underlying under the risk-neutral measure.

2.B Options under the CBOE Pricing Rule

We derive the call option price upper bound under the CBOE margin rule. The pricing formulas for put options can be computed similarly and are not given here. As described in Section 2.2, the margin rule for call options in the CBOE is the piece-wise linear function

$$C(t) = \begin{cases} a_2 S(t) + V(t), & S(t) \le \frac{1}{1+a_1-a_2}K \\ (1+a_1)S(t) - K + V(t), & \frac{1}{1+a_1-a_2}K < S(t) \le K \\ a_1 S(t) + V(t), & S(t) > K \end{cases}$$

We can rewrite equation (2.15) to get

$$V_t'(t) + (r_b - r_d)S(t)V_s'(t) + \frac{1}{2}\sigma^2 S(t)^2 \partial^2 V_{ss}''(t) = r_l V(t) - (r_b - r_l)(C(t) - V(t)). \tag{2.21}$$

Representing equation (2.21) as an expectation value, we obtain an alternative representation of equation (2.17):

$$V(t) = \mathbb{E}_t^{\mathbb{P}_b}\left[e^{-r_l(T-t)}V(T) + \int_t^T e^{-r_l(u-t)}(r_b - r_l)(C(u) - V(u))du\right]. \tag{2.22}$$

[18]Given our assumption that short selling profits earn the lending rate, the solution given by equation (2.20) solves the $\mathscr{M}^+$ for put options for the same reasons that we gave for calls.

Plugging the margin function into equation (2.22) yields

$$\begin{aligned}
V(t) &= \mathbb{E}_t^{\mathbb{P}_b}\left[e^{-r_l(T-t)}V(T)\right] + \mathbb{E}_t^{\mathbb{P}_b}\left[\int_t^T e^{-r_l(u-t)}(r_b-r_l)a_2 S(u)\mathbf{1}_{\{S(u)\le\frac{1}{1+a_1-a_2}K\}}du\right] \\
&\quad +\mathbb{E}_t^{\mathbb{P}_b}\left[\int_t^T e^{-r_l(u-t)}(r_b-r_l)((1+a_1)S(u)-K)\mathbf{1}_{\{\frac{1}{1+a_1-a_2}K<S(u)\le K\}}du\right] \\
&\quad +\mathbb{E}_t^{\mathbb{P}_b}\left[\int_t^T e^{-r_l(u-t)}(r_b-r_l)a_1 S(u)\mathbf{1}_{\{S(u)>K\}}du\right] \\
&= \mathbb{E}_t^{\mathbb{P}_b}\left[e^{-r_l(T-t)}V(T)\right] + a_2(r_b-r_l)\int_t^T e^{-r_l(u-t)}\mathbb{E}_t^{\mathbb{P}_b}\left[S(u)\mathbf{1}_{\{S(u)\le\frac{1}{1+a_1-a_2}K\}}\right]du \\
&\quad +(1+a_1)(r_b-r_l)\int_t^T e^{-r_l(u-t)}\mathbb{E}_t^{\mathbb{P}_b}\left[S(u)\mathbf{1}_{\{\frac{1}{1+a_1-a_2}K<S(u)\le K\}}\right]du \\
&\quad -(r_b-r_l)K\int_t^T e^{-r_l(u-t)}\mathbb{E}_t^{\mathbb{P}_b}\left[\mathbf{1}_{\{\frac{1}{1+a_1-a_2}K<S(u)\le K\}}\right]du \\
&\quad +a_1(r_b-r_l)\int_t^T e^{-r_l(u-t)}\mathbb{E}_t^{\mathbb{P}_b}\left[S(u)\mathbf{1}_{\{S(u)>K\}}\right]du. \qquad (2.23)
\end{aligned}$$

The first term is just the Black–Scholes price under a different measure. To compute the conditional expectations, for notational convenience, we let

$$\begin{aligned}
d_{1,2}(u,t) &= \frac{\ln(\frac{S(t)}{K})+(r_b-r_d\pm\frac{1}{2}\sigma^2)(u-t)}{\sigma\sqrt{u-t}}, \\
d^*_{1,2}(u,t) &= \frac{\ln(\frac{S(t)(1+a_1-a_2)}{K})+(r_b-r_d\pm\frac{1}{2}\sigma^2)(u-t)}{\sigma\sqrt{u-t}}.
\end{aligned}$$

Under the probability measure $\mathbb{P}_b$, we have $dS(t)/S(t) = (r_b-r_d)dt+\sigma dW^b(t)$. Moreover, $W^b(u)-W^b(t)$ is a zero-mean normal variable with variance $u-t$. The conditional expectations can be computed as follows:

$$\begin{aligned}
&\mathbb{E}_t^{\mathbb{P}_b}[S(u)\mathbf{1}_{\{S(u)\le\frac{1}{1+a_1-a_2}K\}}] \\
&= \mathbb{E}_t^{\mathbb{P}_b}\left[S(t)e^{(r_b-r_d-\frac{1}{2}\sigma^2)(u-t)+\sigma(W^b(u)-W^b(t))}\mathbf{1}_{\left\{W^b(u)-W^b(t)\le -d_2^*(u,t)\sqrt{u-t}\right\}}\right] \\
&= \frac{S(t)}{\sqrt{2\pi(u-t)}}e^{(r_b-r_d-\frac{1}{2}\sigma^2)(u-t)}\left(\int_{-\infty}^{-d_2^*(u,t)\sqrt{u-t}} e^{\sigma y}e^{-\frac{y^2}{2(u-t)}}dy\right) \\
&= \frac{S(t)}{\sqrt{2\pi(u-t)}}e^{(r_b-r_d-\frac{1}{2}\sigma^2)(u-t)}\left(\int_{-\infty}^{-d_2^*(u,t)\sqrt{u-t}} e^{-\frac{(y-\sigma(u-t))^2}{2(u-t)}+\frac{1}{2}\sigma^2(u-t)}dy\right) \\
&= S(t)e^{(r_b-r_d)(u-t)}N\Big(-d_1^*(u,t)\Big),
\end{aligned}$$

$$
\begin{aligned}
& \mathbb{E}^{\mathbb{P}_b}\left[S(u)\mathbf{1}_{\{\frac{1}{1+a_1-a_2}K<S(u)\leq K\}}\right] \\
= \;& \mathbb{E}^{\mathbb{P}_b}\left[S(t)e^{(r_b-r_d-\frac{1}{2}\sigma^2)(u-t)+\sigma(W^b(u)-W^b(t))}\mathbf{1}_{\left\{-d_2^*(u,t)\sqrt{u-t}<W^b(u)-W^b(t)\leq -d_2(u,t)\sqrt{u-t}\right\}}\right] \\
= \;& \frac{S(t)}{\sqrt{2\pi(u-t)}}e^{(r_b-r_d-\frac{1}{2}\sigma^2)(u-t)}\left(\int_{-d_2^*(u,t)\sqrt{u-t}}^{-d_2(u,t)\sqrt{u-t}} e^{\sigma y}e^{-\frac{y^2}{2(u-t)}}dy\right) \\
= \;& \frac{S(t)}{\sqrt{2\pi(u-t)}}e^{(r_b-r_d-\frac{1}{2}\sigma^2)(u-t)}\left(\int_{-d_2^*(u,t)\sqrt{u-t}}^{-d_2(u,t)\sqrt{u-t}} e^{-\frac{(y-\sigma(u-t))^2}{2(u-t)}+\frac{1}{2}\sigma^2(u-t)}dy\right) \\
= \;& S(t)e^{(r_b-r_d)(u-t)}\Big(N\big(d_1^*(u,t)\big)-N\big(d_1(u,t)\big)\Big),
\end{aligned}
$$

$$
\begin{aligned}
& \mathbb{E}_t^{\mathbb{P}_b}[\mathbf{1}_{\{\frac{1}{1+a_1-a_2}K<S(u)\leq K\}}] \\
= \;& \mathbb{E}_t^{\mathbb{P}_b}\left[\mathbf{1}_{\{-d_2^*(u,t)\sqrt{u-t}<W^b(u)-W^b(t)\leq -d_2(u,t)\sqrt{u-t}\}}\right] \\
= \;& N\Big(d_2^*(u,t)\Big)-N\Big(d_2(u,t)\Big),
\end{aligned}
$$

$$
\begin{aligned}
& \mathbb{E}_t^{\mathbb{P}_b}[S(u)\mathbf{1}_{\{S(u)>K\}}] \\
= \;& \mathbb{E}\left[S(t)e^{(r_b-r_d-\frac{1}{2}\sigma^2)(u-t)+\sigma(W^b(u)-W^b(t))}\mathbf{1}_{\{W^b(u)-W^b(t)>-d_2(u,t)\sqrt{u-t}\}}\right] \\
= \;& \frac{S(t)}{\sqrt{2\pi(u-t)}}e^{(r_b-r_d-\frac{1}{2}\sigma^2)(u-t)}\left(\int_{-d_2(u,t)\sqrt{u-t}}^{+\infty} e^{\sigma y}e^{-\frac{y^2}{2(u-t)}}dy\right) \\
= \;& \frac{S(t)}{\sqrt{2\pi(u-t)}}e^{(r_b-r_d-\frac{1}{2}\sigma^2)(u-t)}\left(\int_{-d_2(u,t)\sqrt{u-t}}^{+\infty} e^{-\frac{(y-\sigma(u-t))^2}{2(u-t)}+\frac{1}{2}\sigma^2(u-t)}dy\right) \\
= \;& S(t)e^{(r_b-r_d)(u-t)}N\Big(d_1(u,t)\Big).
\end{aligned}
$$

Inserting these expectations into the call option value (2.23) yields

$$
\begin{aligned}
V_{call}(t) \;=\;& S(t)e^{(r_b-r_d-r_l)(T-t)}N(d_1(T,t))-Ke^{-r_l(T-t)}N(d_2(T,t)) \\
& +(r_b-r_l)S(t)\int_t^T e^{(r_b-r_d-r_l)(u-t)}\left(a_2N(-d_1^*(u,t))+a_1N(d_1(u,t))\right)du \\
& +(1+a_1)(r_b-r_l)S(t)\int_t^T e^{(r_b-r_d-r_l)(u-t)}\left(N(d_1^*(u,t))-N(d_1(u,t))\right)du \\
& -(r_b-r_l)K\int_t^T e^{-r_l(u-t)}\left(N(d_2^*(u,t))-N(d_2(u,t))\right)du.
\end{aligned}
$$

2.C Option Pricing with Stochastic Volatility

We derive option price bounds when the underlying asset has stochastic volatility. First, we derive the upper bounds for options. We hedge options with a hedging strategy $U(t)$. The hedging strategy is a five-dimensional process $(\alpha(t), \beta(t), \lambda(t), \delta(t), \gamma(t))$, which captures the holding of the stock, the cash borrowed, the cash deposited, the collateral, and the holding of the contingent claim. We require the strategy to be self-financing. Therefore,

$$dU(t) = \alpha(t)dS(t) + r_b\beta(t)dt + r_l\lambda(t)dt + r_l\delta(t)dt + \gamma(t)dg(t).$$

Plugging the dynamics of $S(t)$ and $g(t)$, we have

$$\begin{aligned} dU(t) &= \gamma(t)\left(g'_t(t) + \mu S(t)g'_s(t) + \frac{1}{2}S^2(t)v(t)g''_{ss}(t)\right)dt \\ &\quad + \left(\kappa(\theta - v(t))g'_v(t) + \frac{1}{2}\xi^2 v(t)g''_{vv}(t) + \xi S(t)v(t)\rho g''_{sv}(t)\right)dt \\ &\quad + \mu\alpha(t)S(t)dt + r_b\beta(t)dt + r_l\lambda(t)dt + r_l\delta(t)dt \\ &\quad + \left(\gamma(t)S(t)\sqrt{v(t)}g'_s(t) + \alpha(t)S(t)\sqrt{v(t)}\right)dW^1(t) + \xi\gamma(t)\sqrt{v(t)}g'_v(t)dW^2(t). \end{aligned}$$

Since the upper bound price equals the value of the hedging strategy, we have $V(t) = U(t)$, and thus $dV(t) = dU(t)$. Applying Ito's formula to $V(t)$ yields

$$\begin{aligned} dV(t) &= \left(V'_t(t) + \mu S(t)V'_s(t) + \frac{1}{2}S(t)^2 v(t)V''_{ss}(t)\right)dt \\ &\quad + \left(\kappa(\theta - v(t))V'_v(t) + \frac{1}{2}\xi^2 v(t)V''_{vv}(t) + \xi S(t)v(t)\rho V''_{sv}(t)\right)dt \\ &\quad + S(t)\sqrt{v(t)}V'_s(t)dW^1(t) + \xi\sqrt{v(t)}V'_v(t)dW^2(t). \end{aligned}$$

Comparing the dynamics of $V(t)$ and $U(t)$, it must hold that

$$\begin{aligned} S(t)\sqrt{v(t)}V'_s(t) &= \gamma(t)S(t)\sqrt{v(t)}g'_s(t) + \alpha(t)S(t)\sqrt{v(t)}, \\ \xi\sqrt{v(t)}V'_v(t) &= \xi\gamma(t)\sqrt{v(t)}g'_v(t). \end{aligned}$$

Solving the equations and using the fact that the additional asset is a variance swap, we obtain

$$\begin{aligned} \gamma(t) &= \frac{V'_v(t)}{g'_v(t)}, \\ \alpha(t) &= V'_s(t) - \frac{V'_v(t)g'_s(t)}{g'_v(t)} = V'_s(t). \end{aligned}$$

Assume that the margin requirement imposed on the whole trading strategy is $C(t)$. Then, $\delta(t) = C(t)$. To investigate the cash flow in the borrowing and lending account, from $V(t) = U(t) = \alpha(t)S(t) + \beta(t) + \lambda(t) + \delta(t) + \gamma(t)g(t)$, we obtain that $\beta(t) + \lambda(t) = V(t) - \alpha(t)S(t) - C(t) - \gamma(t)g(t)$. We need to separate the cases for put options and call options. Since the additional contingent claim is a variance swap, we have $g'_s(t) = 0$. The risk in the underlying asset is hedged by trading the underlying asset. For call options, $\alpha(t) = V'_s(t) > 0$. Since the proceedings of selling call options is always required to be posted as part of the collateral, we have $C(t) \geq V(t)$ and therefore $V(t) - \alpha(t)S(t) - C(t) < 0$. However, $\beta(t) + \lambda(t) = V(t) - \alpha(t)S(t) - C(t) - \gamma(t)g(t)$ can be positive or negative. We cannot determine the sign of this amount. We can approximate it by applying the borrowing rate r_b to this quantity.[19] When this amount is negative, the borrowing rate should be applied. When this amount turns positive, it should earn lending rate. Hence, applying the borrowing rate r_b to this quantity underestimates the cost of replicating. The call option upper bound should be larger than the price derived below. Comparing the drift term of $U(t)$ and $V(t)$ and noting that $g(t)$ is not related to $S(t)$, we derive the following PDE

$$\begin{aligned}
&V'_t(t) + \mu S(t)V'_s(t) + \frac{1}{2}S^2(t)v(t)V''_{ss}(t) + \kappa(\theta - v(t))V'_v(t) + \frac{1}{2}\xi^2 v(t)V''_{vv}(t) + \xi S(t)v(t)\rho V''_{sv}(t) \\
&= \gamma_t \left(g'_t(t) + \kappa(\theta - v(t))g'_v(t) + \frac{1}{2}\xi^2 v(t)g''_{vv}(t)\right) + \alpha(t)\mu S(t) + r_b\beta(t) + r_l C(t) \\
&= \frac{V'_v(t)}{g'_v(t)}\left(g'_t(t) + \kappa(\theta - v(t))g'_v(t) + \frac{1}{2}\xi^2 v(t)g''_{vv}(t)\right) + \mu V'_s(t)S(t) \\
&\qquad + r_b\left(V(t) - V'_s(t)S(t) - C(t) - \frac{V'_v(t)}{g'_v(t)}g(t)\right) + r_l C(t).
\end{aligned}$$

By canceling and rearranging terms, we obtain

$$\begin{aligned}
&\frac{1}{V'_v(t)}\left(V'_t(t) + \frac{1}{2}S^2(t)v(t)V''_{ss}(t) + \kappa(\theta - v(t))V'_v(t) + \frac{1}{2}\xi^2 v(t)V''_{vv}(t) + \xi S(t)v(t)\rho V''_{sv}(t) - r_b V(t)\right) \\
&+ \frac{1}{V'_v(t)}\left(r_b V'_s(t)S(t) + (r_b - r_l)C(t)\right) = \frac{1}{g'_v(t)}\left(g'_t(t) + \kappa(\theta - v(t))g'_v(t) + \frac{1}{2}\xi^2 v(t)g''_{vv}(t) - r_b g(t)\right).
\end{aligned}$$

If we denote the right hand side term the modified variance risk premium and assume it is proportional to $v(t)$, i.e. it equals to $m^* v(t)$, we obtain the following PDE

$$\begin{aligned}
&V'_t(t) + \frac{1}{2}S(t)^2 v(t)V''_{ss}(t) + (\kappa + m^*)\left(\frac{\kappa}{\kappa + m^*}\theta - v(t)\right)V'_v(t) + \frac{1}{2}\xi^2 v(t)V''_{vv}(t) \\
&+ \xi S(t)v(t)\rho V''_{sv}(t) - r_b V(t) + r_b V'_s(t)S(t) + (r_b - r_l)C(t) = 0.
\end{aligned}$$

For put options, $\alpha(t) = V'_s(t) < 0$, therefore the short sell proceed $\alpha(t)S(t)$ is deposited in the cash account and is kept by the broker. We also apply the borrowing rate to the amount $V(t) - C(t) - \gamma(t)g(t)$ as an approximation.[20]Again this approximation underestimates the cost of replicating and thus gives

[19]One can also use the lending rate r_l to approximate the bound. Since using either the borrowing rate or the lending rate would lead to an upper bound which is lower than the actual bound, we use the larger of the two bounds, i.e. the one derived using the borrowing rate, to approximate the upper bound.

[20] Using the lending rate also gives an approximation for the upper bound. But we use the borrowing rate to derive an upper bound that is closer to the actual upper bound.

an upper bound that is lower than the actual upper bound. We then obtain a similar equation

$$V'_t(t) + \frac{1}{2}S(t)^2 v(t) V''_{ss}(t) + (\kappa + m^*)\left(\frac{\kappa}{\kappa + m^*}\theta - v(t)\right) V'_v(t) + \frac{1}{2}\xi^2 v(t) V''_{vv}(t)$$
$$+\xi S(t) v(t) \rho V''_{sv}(t) - r_b V(t) + r_l V'_s(t) S(t) + (r_b - r_l) C(t) = 0.$$

To derive the option price lower bound, we replicate the negative of the options payoffs with the lowest possible capital. Note that the initial capital to hedge the opposite of the option payoff is less than zero. The negative of the minimal initial capital is the lower bound of the option prices. In other words, the highest possible benefit of replicating the opposite of the option payoffs gives the lower bounds. We note that there is no collateral requirement for buying options. Therefore, $\delta(t) = 0$. Similar to solving the upper bound, we have $\gamma(t) = \frac{V'_v(t)}{g'_v(t)}$ and $\alpha(t) = V'_s(t)$. For the borrowing and lending account, $\beta(t) + \lambda(t) = V(t) - \alpha(t)S(t) - C(t) - \gamma(t)g(t)$ incurs the borrowing rate when it is negative and the lending rate when it is positive. For call options, one short sells the underlying to hedge the opposite of the calls, $\alpha(t) = V'_s(t) < 0$. We apply the lending rate to $V(t) - \alpha(t)S(t) - \gamma(t)g(t)$ as an approximation which overestimates the benefit of replicating.[21] Hence we approximate the lower bound for call options by the solution to the following equation

$$V'_t(t) + \frac{1}{2}S(t)^2 v(t) V''_{ss}(t) + (\kappa + m^*)\left(\frac{\kappa}{\kappa + m^*}\theta - v(t)\right) V'_v(t) + \frac{1}{2}\xi^2 v(t) V''_{vv}(t)$$
$$+\xi S(t) v(t) \rho V''_{sv}(t) - r_l V(t) + r_l V'_s(t) S(t) = 0.$$

To replicate the opposite of puts, $\alpha(t) = V'_s(t) > 0$. We apply the borrowing rate to $\beta(t) + \lambda(t) = V(t) - \alpha(t)S(t) - \gamma(t)g(t)$.[22] The lower bound for puts then solves

$$V'_t(t) + \frac{1}{2}S(t)^2 v(t) V''_{ss}(t) + (\kappa + m^*)\left(\frac{\kappa}{\kappa + m^*}\theta - v(t)\right) V'_v(t) + \frac{1}{2}\xi^2 v(t) V''_{vv}(t)$$
$$+\xi S(t) v(t) \rho V''_{sv}(t) - r_b V(t) + r_b V'_s(t) S(t) = 0.$$

The actual lower bounds would be below the lower bounds we derived above. The upper and lower bounds can be solved numerically using finite difference methods.

The upper and lower bounds are compared with the Heston price derived from the following equation

$$V'_t(t) + \frac{1}{2}S^2(t) v(t) V''_{ss}(t) + (\kappa + m^*)\left(\frac{\kappa}{\kappa + m^*}\theta - v(t)\right) V'_v(t) + \frac{1}{2}\xi^2 v(t) V''_{vv}(t)$$
$$+\xi S(t) v(t) \rho V''_{sv}(t) - r_l V(t) + r_l V'_s(t) S(t) = 0.$$

[21] Using the borrowing rate gives a lower bound which is higher than the lower bound derived from the lending rate. Thus, the lending rate is used.

[22] The borrowing rate is chosen here, as it yields a lower lower bound.

	Call options			Put options		
	$\exp(\Pi_0)$	Π	Adjusted R^2	$\exp(\Pi_0)$	Π	Adjusted R^2
Short-term	0.184	-3.381	0.730	0.191	-2.923	0.712
	(-62.210)	(-32.104)		(-59.085)	(-29.852)	
Medium-term	0.186	-2.105	0.855	0.192	-1.957	0.920
	(-68.913)	(-37.549)		(-67.193)	(-43.247)	

TABLE 2.1: *Regression results for obtaining the empirical slopes for short-term and medium-term options*

The table displays the results for the regression (2.8) of implied volatility on moneyness for call and put options with t-statistics in parentheses. We ran the regression for each week of our sample period from January 2002 to August 2010 for a total of 449 weeks. The term $\exp(\Pi_0)$ represents the implied volatility for at-the-money options. The reported coefficients and adjusted R^2 are time averages over all 449 weeks. The t-statistics are the time-series average of the weekly estimates divided by the standard deviation of the average adjusted for first-order autocorrelation (BKM (2003)). Short-term options are those with maturities less than 60 days. Medium-term options have expirations between 60 to 150 days.

	Short-term				Medium-term			
	only lag	strategy margin	portfolio short	minimum portfolio	only lag	strategy margin	portfolio short	minimum portfolio
	Panel A: Call options							
$\Delta\Pi_t^{\text{model}}$		3.253	2.569	2.592		1.763	1.536	1.551
		(8.844)	(9.540)	(9.562)		(13.704)	(15.304)	(15.341)
$\Delta\Pi_{t-1}$	-0.238	-0.225	-0.222	-0.222	-0.286	-0.295	-0.288	-0.288
	(-4.968)	(-5.983)	(-6.170)	(-6.178)	(-6.677)	(-6.469)	(-6.450)	(-6.454)
Adjusted R^2	0.054	0.227	0.257	0.258	0.080	0.408	0.421	0.421
	Panel B: Put options							
$\Delta\Pi_t^{\text{model}}$		6.603	10.679	5.066		3.988	8.409	4.128
		(9.168)	(5.244)	(7.617)		(13.052)	(9.756)	(12.600)
$\Delta\Pi_{t-1}$	-0.312	-0.317	-0.306	-0.314	-0.195	-0.218	-0.163	-0.165
	(-5.201)	(-6.009)	(-5.578)	(-5.983)	(-7.131)	(-6.251)	(-5.543)	(-5.385)
Adjusted R^2	0.095	0.238	0.200	0.238	0.036	0.463	0.277	0.357

TABLE 2.2: *Regression results for changes of empirical IV slopes on changes of model-implied slopes.*

The table reports the estimated coefficients from regressing the differences of empirical IV slope on the lagged differences and also on differences of model-implied slopes. We give corresponding t-statistics in parentheses. Panel A shows the results for both short-term and medium-term call options. Panel B reports the results for short-term and medium-term put options. For each option category, we report the results of the regression using the lagged dependent variable alone and also for the combined regression using three margin rules, namely strategy margins for a naked short sale, the portfolio margins for a naked short sale, and minimum portfolio margin requirements. The standard errors used to compute the t-statistics are the Newey–West estimates with a lag length of 5.

	Short-term				Medium-term			
	only controls	strategy margin	portfolio short	minimum portfolio	only controls	strategy margin	portfolio short	minimum portfolio
	Panel A: Call options							
$\Delta\Pi_t^{\text{model}}$		3.276	2.586	2.609		1.706	1.496	1.510
		(8.839)	(9.518)	(9.541)		(12.017)	(13.597)	(13.638)
$\Delta\text{Skewness}_t$	0.108	0.177	0.194	0.194	-0.222	-0.034	0.012	0.014
	(0.371)	(0.695)	(0.788)	(0.790)	(-1.585)	(-0.300)	(0.109)	(0.128)
$\Delta\text{Kurtosis}_t$	0.007	0.023	0.025	0.177	-0.078	-0.019	-0.012	-0.011
	(0.203)	(0.743)	(0.824)	(0.824)	(-2.544)	(-0.793)	(-0.526)	(-0.514)
$\Delta\Pi_{t-1}$	-0.237	-0.227	-0.224	-0.224	-0.300	-0.300	-0.291	-0.291
	(-4.850)	(-5.955)	(-6.155)	(-6.163)	(-6.455)	(-6.258)	(-6.205)	(-6.205)
Adjusted R^2	0.051	0.225	0.255	0.207	0.126	0.409	0.422	0.422
	Panel B: Put options							
$\Delta\Pi_t^{\text{model}}$		6.684	10.723	5.106		3.889	7.907	3.955
		(9.056)	(5.295)	(7.641)		(11.549)	(8.235)	(12.178)
$\Delta\text{Skewness}_t$	0.228	0.344	0.270	0.317	-0.300	-0.055	-0.101	-0.070
	(0.942)	(1.725)	(1.409)	(1.651)	(-0.168)	(-0.561)	(-0.851)	(-0.644)
$\Delta\text{Kurtosis}_t$	0.021	0.038	0.025	0.034	-0.075	-0.016	-0.033	-0.025
	(0.869)	(1.839)	(1.248)	(1.695)	(-3.139)	(-0.968)	(-1.453)	(-1.222)
$\Delta\Pi_{t-1}$	-0.313	-0.314	-0.307	-0.312	-0.172	-0.221	-0.172	-0.171
	(-5.199)	(-5.894)	(-5.575)	(-5.885)	(-5.667)	(-6.243)	(-5.667)	(-5.566)
Adjusted R^2	0.093	0.239	0.199	0.239	0.089	0.463	0.288	0.335

TABLE 2.3: *Regression results for changes of empirical IV slopes on changes of risk-neutral parameters.*

The table reports the estimated coefficients for regressions explaining the difference of slopes using the difference of risk-neutral parameters. Panel A shows the results for call options and Panel B for put options. We analyze short-term and medium-term options separately. Column *only controls* shows the regression where only control variables are employed. We also run combined regressions using implied slopes. Results for regressions incorporating slopes derived from each type of margin rule are shown in the column labeled according to the margin rule. The standard errors used to compute the t-statistics are the Newey–West estimates with a lag length of 5.

	Short-term				Medium-term			
	only controls	strategy margin	portfolio short	minimum portfolio	only controls	strategy margin	portfolio short	minimum portfolio
	Panel A: Call options							
$\Delta\Pi_t^{\text{model}}$		3.374	2.627	2.648		1.725	1.484	1.495
		(8.709)	(9.080)	(9.093)		(10.224)	(13.705)	(13.752)
$\Delta\text{LiborOIS}_t$	-0.207	0.524	0.536	0.530	-0.081	0.261	0.231	0.226
	(-0.714)	(1.929)	(1.210)	(2.138)	(-0.649)	(1.676)	(2.129)	(2.114)
ΔVIX_t	0.075	-0.023	-0.021	-0.021	0.044	0.004	0.007	0.007
	(3.500)	(-1.837)	(-1.070)	(-1.673)	(4.472)	(0.941)	(1.584)	(1.681)
$\Delta\text{IndexReturn}_t$	-0.277	-0.428	-0.522	-0.525	0.020	-0.154	-0.172	-0.173
	(-0.289)	(-0.452)	(-0.409)	(-0.555)	(0.060)	(-0.469)	(-0.522)	(-0.527)
$\Delta\text{RelativeDemand}_t$	-0.356	-0.273	-0.237	-0.237	0.004	0.010	0.003	0.003
	(-2.700)	(-2.215)	(-2.285)	(-1.954)	(0.075)	(0.227)	(0.089)	(0.084)
$\Delta\Pi_{t-1}$	-0.221	-0.218	-0.217	-0.217	-0.286	-0.299	-0.291	-0.291
	(-4.956)	(-5.754)	(-5.959)	(-5.307)	(-7.301)	(-6.720)	(-6.712)	(-6.717)
Adjusted R^2	0.116	0.235	0.262	0.263	0.159	0.409	0.424	0.424
	Panel B: Put options							
$\Delta\Pi_t^{\text{model}}$		5.995	9.625	4.600		3.755	8.276	3.903
		(7.229)	(4.619)	(6.137)		(11.895)	(9.865)	(11.962)
$\Delta\text{LiborOIS}_t$	-0.262	0.121	-0.201	0.050	-0.112	0.015	-0.142	-0.064
	(-0.808)	(0.505)	(-0.806)	(0.238)	(-0.997)	(0.230)	(-1.357)	(-0.747)
ΔVIX_t	0.083	0.026	0.062	0.041	0.036	0.017	0.034	0.028
	(3.430)	(2.282)	(3.516)	(3.054)	(4.298)	(3.627)	(4.538)	(4.200)
$\Delta\text{IndexReturn}_t$	-0.986	-1.242	-1.427	-1.448	0.040	-0.159	-0.019	-0.069
	(-1.156)	(-1.563)	(-1.749)	(-1.732)	(0.128)	(-0.522)	(-0.064)	(-0.221)
$\Delta\text{RelativeDemand}_t$	-0.029	0.063	0.051	0.089	0.023	0.025	0.006	0.006
	(-0.231)	(0.535)	(0.431)	(0.737)	(0.584)	(0.883)	(0.201)	(0.220)
$\Delta\Pi_{t-1}$	-0.322	-0.323	-0.316	-0.322	-0.204	-0.225	-0.170	-0.174
	(-5.634)	(-6.246)	(-5.952)	(-6.316)	(-7.160)	(-6.239)	(-5.940)	(-5.574)
Adjusted R^2	0.149	0.241	0.230	0.252	0.127	0.484	0.362	0.412

TABLE 2.4: *Regression results for changes of empirical IV slopes using the second set of control variables.*

The table shows the estimated coefficients from regressing the changes of empirical slopes on changes of the second set control variables. Panel A reports the results for calls, Panel B those for puts. We analyze short-term and medium-term options separately. Column *only controls* shows the regression where only control variables are employed. We also run combined regressions using implied slopes. Results for regressions including slopes derived from each type of margin rule are shown in the column labeled according to the margin rule. The standard errors used to compute the t-statistics are the Newey–West estimates with a lag length of 5.

Chapter 3

How Index Futures and ETFs Increase Stock Return Correlations

Abstract

This paper examines whether increased trading activity in index futures and exchange traded funds (ETFs) is associated with higher equity return correlations. We build a simple model to analyze how demand shocks for index ETFs and futures are transmitted to the underlying stocks through arbitrage. Our model predicts that demand shocks to ETFs and the futures market lead to a stronger price comovement not only for index stocks but also for non-index stocks. Moreover, demand shocks to index ETFs have a higher impact on stock return correlations than demand shocks to futures. We confirm the model predictions by studying the average pairwise correlation of S&P 500 stocks after the inception of S&P 500 futures. Controlling for several factors, we show that trading activity in futures and ETFs explains the time variation of the average S&P 500 stock return correlation, with ETFs exhibiting a significantly stronger explanatory power. An examination of the relationship between current and lagged returns suggests that at least some of the return comovement is excessive. Furthermore, we confirm that futures and ETF trading activity is also associated with higher return correlations among non-index stocks.

We thank Henrik Hasseltoft, Markus Leippold, Per Östberg, Tatjana Puhan, Alexandre Ziegler for comments and discussions and participants in the Brown bag seminar at the University of Zurich for feedback. Financial support by the Swiss Finance Institute (SFI) is gratefully acknowledged.

3.1 Introduction

Covariances between asset returns are a key variable in risk management, portfolio selection and tests of asset pricing models. The covariance between assets can change due to changes in volatilities or changes in correlations. Both volatilities and correlations between assets vary substantially over time. Previous work documents that correlations between international equity returns change over time (see, for example, Longin and Solnik (2001), Goetzmann, Li, and Rouwenhorst (2005) and Ang and Bekaert (2002)). There is also evidence that correlations among stock returns in the US vary considerably over time (see, for example, Campbell, Lettau, Malkiel, and Xu (2001)). Typically, correlations rise in periods of financial crises and high volatility.

In traditional theory, asset returns are correlated because changes in their fundamentals are correlated. However, the asset correlations observed in the data sometimes exceed the level predicted by fundamentals. Pindyck and Rotemberg (1993) select stocks from unrelated lines of business and find excess comovement in returns. Moreover, many researchers find that the addition and removal of a stock from the index affects its correlation with the index, although the addition and removal itself might not convey new information above and beyond stock fundamentals (Vijh (1994), Barberis, Shleifer, and Wurgler (2005), Greenwood (2005)).

This paper investigates the impact of trading activity in index ETFs and futures on asset return correlations. Index or sector ETFs and futures have gained increasing popularity among investors. However, the consequences of this increase in trading opportunities for market correlation are not well understood at a theoretical or empirical level. Increased trading in ETFs and futures might increase stock return correlations due to the no-arbitrage relation between the derivative and the underlying assets. The idea is simple: when a demand shock hits the index derivative, the price of the derivative will move away from the price implied by the underlying basket. However, the no-arbitrage relation between the index derivative and the underlying baskets leads arbitrageurs to attempt to exploit the mispricing by buying the under-priced asset and selling the over-priced asset. Because this arbitrage involves simultaneous trades in the index constituents, correlations among stocks are likely to increase.

To demonstrate how trading activity in ETFs and futures affects the correlations among assets, we build a simple two-period model. Investors trade in the first period and consume the liquidation value of all assets in the second. The economy consists of three types of assets: a riskless asset, N risky assets and index derivatives, i.e., futures and physically replicating ETFs. We introduce quadratic transaction costs in our model so that index derivatives are not redundant and arbitrage is costly. The model features three types of investors, namely stock investors, index traders and arbitrageurs. We show that given an exogenous demand shock from index traders, arbitrageurs, as the only type of investors present in both the derivative and stock markets, provide the derivatives to index investors while hedging themselves by taking positions in the stock market. Thus, demand shocks from index derivatives are transmitted to the stock market through arbitrageurs. In equilibrium, demand shocks from index traders also play a role in determining stock prices. Hence, in addition to fundamental correlations, stock prices can also be correlated through exposures to demand shocks. The model predicts that demand shocks to physically replicated ETFs have a higher impact on correlations than demand shocks to futures, a prediction that is borne out in the data.

To test for a relationship between index trading and stock return correlation in the data, we investigate the impact of S&P 500 futures, E-mini futures and S&P 500 ETFs on average stock return correlations. Our data range from January 1982 to December 2012. We use the monthly value-weighted return correlations among S&P 500 stocks as well as non-S&P 500 stocks as the dependent variables. As a proxy for the demand shocks to ETFs (futures), we use the ratio of ETFs (futures) dollar trading volume and aggregate S&P 500 stock dollar trading volume. Consistent with the model predictions, we find that the demand shock proxy significantly explains the time variation of both S&P 500 and non-S&P 500 stock correlations. The R^2 increases substantially when the demand shock proxy is added to the regression. The results remain robust when controlling for the current month return, realized volatility and many other factors. Notably, in line with the model predictions, we find that ETF trading activity has a larger impact on correlations than futures trading activity. We also compare the impact of index demand shocks on the correlations of index and non-index stocks by pooling all correlations in one regression. In line with our intuition, index demand shocks have a larger impact on index stocks. As trading activity in index products increases, the average stock correlation increases. A natural question is whether such an increase in correlations is excessive or whether it reflects information that is transmitted from the index derivatives market to the stock market. To answer this question, we test whether higher levels of stock market return reversals are associated with higher correlations. Return reversals measure how the daily returns are linked with the lagged daily returns. A stronger negative relationship between today's and the previous day's returns indicates more return reversals. By value weighting the return reversals of all index stocks, we are able to derive a return reversal measure for the index. We test the excessiveness of the correlation by regressing the return reversals on the correlations. The regression results confirm that a higher correlation is significantly associated with larger return reversals. Hence, at least part of the correlation is excessive.

Our paper is related to three strands of the literature. The first strand of research investigates whether correlation risk is priced. Krishnan, Petkova, and Ritchken (2009) document that the time-varying correlation between individual stocks carries a significant price of risk in the cross-section of stock returns. Investors pay a premium for securities that perform well in high correlation regimes. In the options market, Driessen, Maenhout, and Vilkov (2009) show that incorporating a stochastic correlation between asset returns can explain the steepness of the implied volatility curves of index options. Index options are sold at a premium compared to equity options because they can be used to hedge correlation risk. Schürhoff and Ziegler (2011) decompose total variance risk into systematic and idiosyncratic variance and find that both carry risk premia but with different signs. Correlation risk premia are a combination of the risk premia on systematic and idiosyncratic variances.

Our paper is also related to the literature that analyzes the impact of index product trading on the underlying equity market. Papers that study futures trading activity and its impact on stock market volatility find inconsistent results (see, for example, Edwards (1988) and Bessembinder and Seguin (1992)). The empirical evidence on the link between equity futures trading and an increase in stock market volatility is mixed. For ETFs, Ben-David, Franzoni, and Moussawi (2012) show that shocks to ETF prices are passed down to the underlying securities via the arbitrage between the ETF and underlying stocks. They find that ETFs increase the volatility of the underlying securities.

Our paper is most closely related to the literature that investigates the link between ETFs and the correlation between their constituents. Da and Shive (2012) and Staer (2012) study empirically whether an increase in the trading volume of an ETF leads to an increase in the correlations of the ETF component stocks. Staer (2012) studies this relationship using high-frequency data, while Da and Shive (2012) test this hypothesis using daily data. Both papers conclude that due to the trading pressure induced by arbitrageurs, trading activity in ETFs indeed results in higher stock return correlations.

The main innovation of our paper compared to the literature is three-fold. Firstly, we build a theoretical model that allows us to obtain a number of predictions that we are also able to confirm in the data. Other papers only provide empirical evidence on the impact of ETFs on stock volatility and correlations. Secondly, we consider both futures and ETFs and compare their impact on correlations. To our knowledge, futures, as another popular exchange-traded index product that has a much larger trading volume and a much longer history than ETFs, have not been considered to date in the context of return correlations. Furthermore, our model predicts and we confirm in the data that demand shocks coming from physically replicating ETFs have a higher impact than shocks associated with futures trading. Thirdly, our model implies that index demand shocks initiated from index derivatives might not only increase return correlations for index stocks but also induce higher correlations for non-index stocks. Previous work does not consider the impact from trading activity in index products on non-index stocks.

The remainder of this paper is organized as follows. Section 3.2 provides an overview of the futures and ETF markets. Section 3.3 presents a simple equilibrium model demonstrating how demand shocks to index products affect correlations between the underlying stocks. Section 3.4 describes our empirical hypotheses, the data and the primary results of our empirical tests. Section 3.5 concludes the paper.

3.2 An Overview of Futures and ETF Markets

3.2.1 Description of the Instruments

S&P 500 index futures were introduced by the Chicago Mercantile Exchange (CME) on April 21, 1982. One futures contract was 500 times the value of the S&P 500 stock index. The CME reduced the size of its S&P 500 futures contract by reducing the multiplier from 500 to 250 on November 3, 1997. Futures use the open outcry pit trading method during main market hours. At maturity, there is no physical delivery of the underlying asset. Futures are settled in cash to the spot value of the index. S&P 500 futures are listed in a quarterly cycle, with contracts expiring in March, June, September, and December.

E-mini futures were launched on September 9, 1997, because the contract size of the S&P 500 futures contract was too large for small investors. As the name suggests, E-mini futures have a much smaller contract size than legacy S&P 500 index futures. One E-mini S&P 500 future contract is 50 times the value of the S&P 500 stock index. Like other CME stock index futures, E-mini futures are cash settled. CME E-mini contracts are offered exclusively on the CME Globex electronic trading platform and can be traded almost 24 hours per day.

ETFs, as another important type of index derivative, are created by placing assets or total return swaps into a trust. The trust then issues certificates or fund shares that are listed on an exchange. Unlike standard equity mutual funds, ETFs trade like individual stocks and can be bought or sold throughout the trading day, not just at the closing price of the day. Stock index ETF share prices typically represent a certain fraction of their underlying index. For example, the SPDR S&P 500 (SPY), the largest ETF in the market, is valued at approximately one-tenth the value of the index. There are two other ETFs that also track the S&P 500 index, namely the iShares S&P 500 Index (IVV) and the Vanguard 500 Index Fund (VOO).

There are two ways in which an ETF tracks an index or a sector. The first (and most common) method is known as physical replication, where the fund owns the securities represented by the index, providing investors with actual ownership of the securities. Almost all ETFs that track an index or a sector in the United States are physical ETFs.[1] While some funds hold the stocks in the weighting defined by the index, called full replication, others own a proportion of shares in the underlying index to increase efficiency and lower cost, an approach referred to as optimized replication. Optimized replication is usually used for very broad-based indices such as the Russell 1000 or Russell 2000. In our study, the three main S&P 500 index ETFs, i.e., SPY, IVV and VOO hold index constituent stocks with weights that are very close

[1] The physical replication of ETFs can be confirmed by checking the asset holdings of equity ETFs in the Thomson Reuters Mutual Fund Holdings data. The ETF database provides easy access to the holding of top ETFs and thereby confirms physical replication (http://etfdb.com/compare/market-cap/). Vanguard mentions that "nearly all ETF products in the United States are physical ETFs". iShares, the largest ETF provider, writes in its investors' guide that "Indeed, almost 100% of currently offered iShares ETFs are physically-replicated." SPDR also mentions that "To date, all State Street Global Advisors' (SSgA) SPDR ETFs across the globe are supported by the physical replication model using either full or optimised replication."

to the weights in the S&P 500. These three ETFs can be considered to be fully replicating the index. The second method for constructing an ETF is known as synthetic (or swap-based) replication. These funds use swaps and other derivative products to obtain market access. Synthetic ETFs are more popular in Europe. They are more commonly used in real estate, commodities and money markets. Because the focus of our study is the US equity market, we can safely assume physical replication.

3.2.2 Growth in ETFs and Futures

Trading activity in S&P 500 stocks and derivatives has increased over time. Panel A of Figure 3.1 shows that over the past three decades, trading in S&P 500 stocks as measured by aggregate dollar trading volume has expanded almost exponentially.

Accompanying the growth of stock trading is an increase in futures and ETF trading. At first glance, ETFs and futures appear to be highly redundant and incapable of significantly expanding investors' opportunity sets. However, these products are a more cost-effective way of gaining access to an entire market than buying the individual index constituents. Therefore, in terms of dollar trading volume, they are highly successful.

Since the inception of futures contracts in 1982, floor-based futures have been a popular venue for index trading. After the introduction of E-mini futures, trading in traditional futures has trended down as part of the trading has been replaced by trading in E-mini futures. E-mini stock index futures debuted in 1997 and have become the fastest-growing futures products in history. Studies of comparative market quality by Domowitz and Steil (1999) find that electronic markets tend to offer liquidity similar to that of floor markets but at a lower cost. These cost advantages might be one of the reasons, aside from the flexible trading hours and smaller contract size, that E-mini futures have replaced floor-based futures to be the most traded index product. The sum of total futures trading volumes (legacy plus E-mini) remains high, at a level that exceeds the total dollar trading volume in S&P 500 stocks.

The first ETFs were introduced in the early 1990s, but only achieved good market penetration in the mid- to late-90s. In recent years, ETFs have experienced remarkable growth over a relatively short period. In Panel A of Figure 3.1, we report the total dollar trading volume of three ETFs that track the S&P 500 index, i.e., SPY, IVV and VOO.

To compare futures and ETF trading activity with the underlying cash market, Panel B of Figure 3.1 reports the ratio of the dollar trading volume of various products to the aggregate trading volume of S&P 500 stocks. Originally, traditional futures had a much larger trading volume than the cash market. After the inception of E-mini futures, legacy futures lost their popularity. E-mini futures trading volume is 1.5 times the volume on the cash market. Hence, the total volume of legacy futures and E-mini futures usually exceeds that of S&P 500 stocks and is sometimes twice as large. ETFs are not traded as often as their underlying stocks. Their dollar trading volume is generally around 20% of the total dollar trading volume of the under-

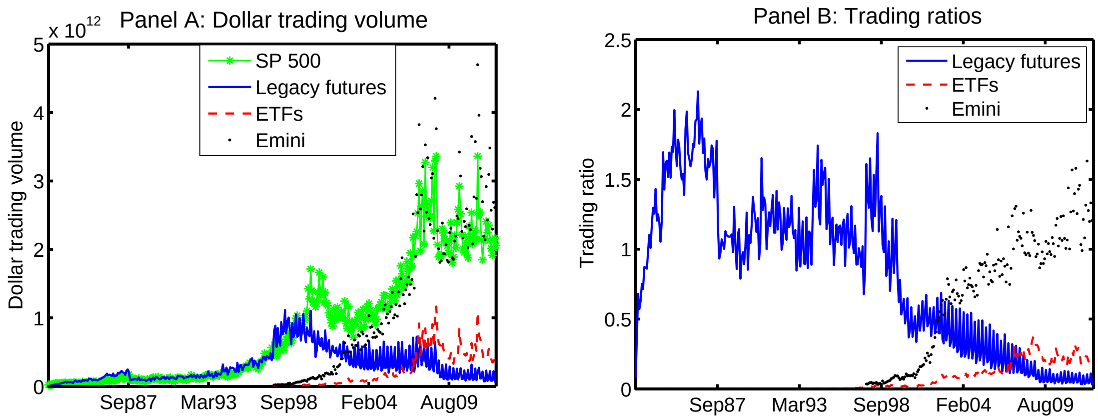

FIGURE 3.1: *Trading activity in S&P 500 stocks, ETFs, legacy futures and E-mini futures over time.* Panel A depicts the monthly dollar trading volume of S&P 500 stocks, futures and ETFs over time. S&P 500 dollar volume is the aggregate dollar volume of S&P 500 index constituents. ETF volume is the sum of SPY, IVV and VOO dollar trading volume. Futures volume is the sum of S&P 500 futures volume across all maturities. Panel B shows the ratio of the dollar trading volume of various products to the aggregate trading volume of S&P 500 stocks.

lying stocks.

3.2.3 Why ETFs and Futures Might Increase Stock Return Correlations

ETFs and futures, as derivatives based on an index, must satisfy a no-arbitrage relationship with the index. When the prices of ETFs and futures move away from the prices implied by the cash market, arbitrageurs can buy the over-priced and sell the under-priced to make a profit. As reaping the arbitrage benefit involves the simultaneous trading of all index constituents, correlations of stocks returns are likely to increase.

Of course, futures and ETFs are not the only instruments whose value depends on that of the underlying basket. Mutual funds and options are also often index-based. Although a large demand shock to index options might also lead to arbitrage between the options and the underlying stocks, the no-arbitrage relation between the index and index options is more difficult to maintain because option prices also depend on (stochastic) volatility. Index mutual funds are another important way to invest in the index, and their capitalization by far exceeds that of ETFs. However, index mutual funds lack a liquid secondary market and can be traded only at the close of the day. Cash inflows and outflows to the mutual fund are likely to have an impact on the correlation of the underlying stocks. Da and Shive (2012) find that mutual fund holdings also explain part of stock return correlations, but their coefficient is only one third that of ETF holdings. This is evidence that ETFs have a higher impact on correlations than mutual funds. In this paper, we focus on the impact of index products that have a liquid secondary market; thus, we do not include index mutual funds in our analysis.

3.3 Model

Our formal analysis is conducted using a simple two-period model that allows us to derive a unique equilibrium. Investors trade at $t = 1$ and then consume the liquidation value of all assets at $t = 2$. The setup we use builds on the work of Fremault (1991), who solves for equilibrium with competitive stock and futures markets. In her paper, there are no demand shocks for the derivative; as a result, she does not analyze how demand shocks on the derivative market affect stock return correlations. Moreover, she does not examine the differences between futures and ETFs.

3.3.1 Financial Assets

The economy consists of three types of assets. There is a riskless asset that has a perfectly elastic supply with a gross rate of return normalized to 1. The economy also contains a fixed supply $\boldsymbol{Q}$ of N risky assets. The third type of assets are derivatives written on an index. The weight of the stocks in the index is $\boldsymbol{b}$, with $\boldsymbol{b}^T\mathbf{1} = 1$. In line with market practice, only a subset of the stocks is included in the index. We assume that the first k stocks out of N are included in the index. Obviously, $b_i \neq 0$ for stock i with $1 \leq i \leq k$. The index itself cannot be traded, but the derivatives written on it can. We consider two types of derivatives written on the index, i.e., a futures contract and an ETF.[2] Both the futures and the ETF are assumed to be in zero net supply.[3] The payoffs of the two derivatives are different. The futures contract is cash settled; its payoff is equal to the difference between the index value at $t = 2$ and the futures price at $t = 1$.[4] The ETF in our model is assumed to physically replicate the index. Therefore, it is a basket of stocks. Trading in the ETF is identical to trading a basket of stocks with weights equal to the index weight $\boldsymbol{b}$. At $t = 2$, the ETF gives the exact same payoff as holding the index.

Trading stocks, ETFs and futures contracts is not costless in our model. The transaction costs can be interpreted as the commissions and the cost of immediacy. Although commissions are quite low for large institutional investors, the cost of immediacy can be sizable even for institutional investors. Without transaction costs, index derivatives are redundant because the payoffs of these securities can be replicated easily and costlessly by investors. In a market with frictions such as trading costs, the existence of ETFs and the futures market allows traders to generate larger order flows at lower costs. To obtain a closed-form solution, we as-

[2] It is worth noting that although we assume that the futures and ETFs are based on a broad index, the same results can be applied to sector-based products such as sector ETFs and futures.

[3] This assumption is not crucial to the model but is imposed for simplicity. To allow for non-zero net supply, we can assume that liquidity providers have a non-zero initial endowment in ETFs and futures.

[4] Because our model has only two periods, there is no mark-to-market between the time the position is entered and the final settlement. The futures contract resembles a forward contract in our setting.

sume quadratic transaction costs. Quadratic transaction costs are not unreasonable and are assumed in many papers for analytical convenience (See Fremault (1991) and Garleanu and Pedersen (2013), for example). In our model, trading costs for stocks are captured in a matrix $\boldsymbol{C}$. To trade $\boldsymbol{X}$ stocks, the associated trading costs are $\frac{1}{2}\boldsymbol{X}^T\boldsymbol{C}\boldsymbol{X}$. The non-diagonal elements of $\boldsymbol{C}$ measure how assets' liquidity is linked. We denote the futures trading cost coefficient by c_f; to trade y futures, the cost is $\frac{1}{2}c_f y^2$. Similarly, denoting the ETF trading cost coefficient by c_{etf}, it costs $\frac{1}{2}c_{etf}z^2$ to trade z ETFs.

3.3.2 Investors and Beliefs

We model three types of investors: stock investors, index traders and liquidity providers. We assume N_s identical stock investors who are restricted to trade only in stocks and the riskless asset. Before trading starts at $t = 1$, they receive an equal endowment of $\frac{\boldsymbol{Q}}{N_s}$ stocks and B_0^S riskless assets. Index traders are assumed to generate exogenous demand shocks for index products. We let q_f denote the exogenous demand for futures and q_{etf} the exogenous demand for ETFs. Demand shocks are taken as given. Finally, there are liquidity providers (arbitrageurs) that trade in both the ETF (futures) market and the stock market. There are N_1 identical futures providers and N_2 identical ETF providers that act competitively in that they do not take into account the price impact of their trades. As the only investors present in both the derivative and the stock markets, liquidity providers not only reap arbitrage profits but also provide liquidity to accommodate exogenous demand shocks. Liquidity providers do not have an initial endowment in risky assets. We assume that each futures and ETF provider has an initial endowment B_0^1 and B_0^2 in the riskless asset, respectively. The market structure is illustrated in Figure 3.3.

Stock investors and liquidity providers have identical prior beliefs about stocks' payoff $\boldsymbol{f}$ in the second period, where $\boldsymbol{f} \sim N(\boldsymbol{\mu}, \boldsymbol{\Sigma})$. At $t = 1$, they observe an unbiased public signal $\boldsymbol{\eta} = \boldsymbol{f} + \boldsymbol{e}_\eta$, where $\boldsymbol{e}_\eta \sim N(\boldsymbol{0}, \boldsymbol{\Sigma}_\eta)$. Aside from the public signal, futures and ETF demand shocks might also contain some information about the payoff in the second period. Specifically, we assume that the futures demand shock is given by $q_f = \boldsymbol{\tau}_1\boldsymbol{f} + e_f$, where $e_f \sim N(0, \sigma_f^2)$, and the ETF demand shock by $q_{etf} = \boldsymbol{\tau}_2\boldsymbol{f} + e_{etf}$, where $e_{etf} \sim N(0, \sigma_{etf}^2)$. We further assume that $\boldsymbol{\tau}_1$ and $\boldsymbol{\tau}_2$ are known to all agents, in particular to stock investors and liquidity providers. Obviously, when $\boldsymbol{\tau}_1$ and $\boldsymbol{\tau}_2$ are zero, demand shocks are pure noise. When $\boldsymbol{\tau}_1$ and $\boldsymbol{\tau}_2$ are non-zero, however, index traders possess superior information about future returns. Hence, demand shocks provide valuable information about the payoff in the second period. Stock investors and liquidity providers combine all of the information observed in the first period

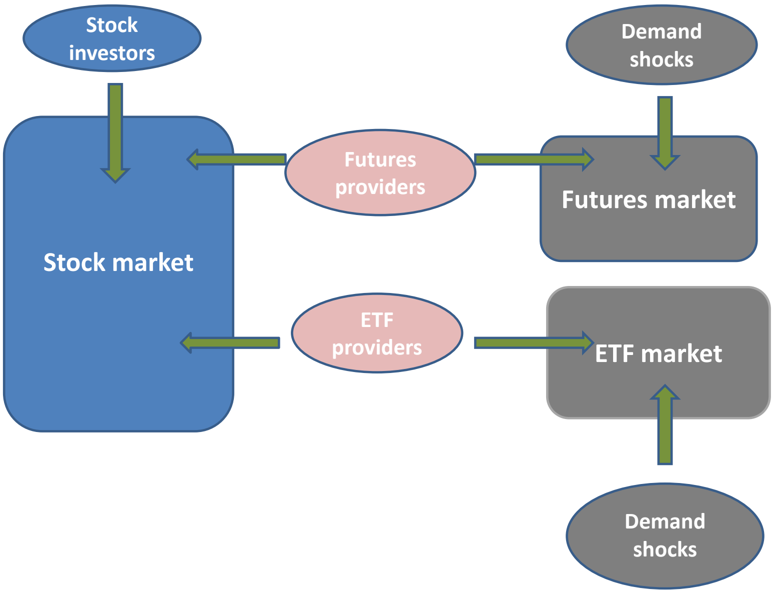

FIGURE 3.2: *Market structure of the model.*
This figure shows the structure of the model. There are three markets: the stock market, the ETF market and the futures market. There are three types of investors. Index traders, assumed to be exogenous, generate demand shocks to the index derivatives market. Stock investors participate only in the stock market. Futures and ETF liquidity providers (arbitrageurs) trade in both the stock market and the derivatives market.

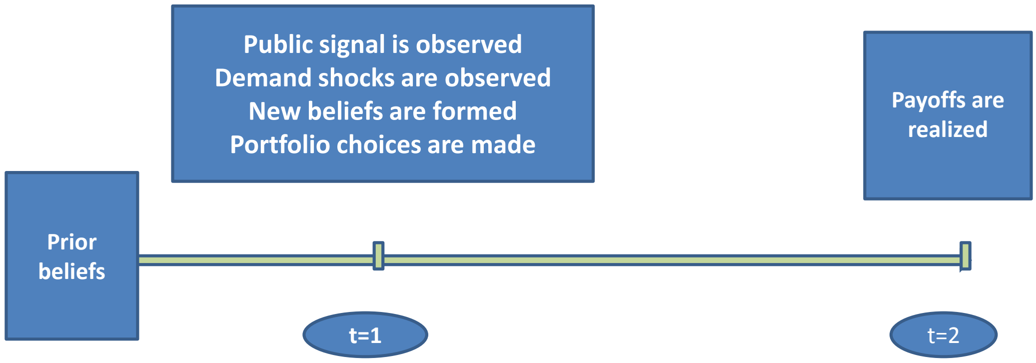

FIGURE 3.3: *Time line of the model.*
This figure shows the sequence of events in the model. Before trading starts, investors have prior beliefs about asset payoffs in the second period. In period 1, investors observe both the public signal and the demand shocks, from which a new belief about the payoff in period 2 is formed. Investors then build their portfolio based on their new beliefs. The equilibrium is reached in period 1, while the payoffs are realized in period 2.

to form posterior beliefs. Applying Bayes' rule, the posterior belief is $\boldsymbol{f} \sim N(\hat{\boldsymbol{\mu}}, \hat{\boldsymbol{\Sigma}})$, where

$$\begin{aligned}\hat{\boldsymbol{\mu}} &= \hat{\boldsymbol{\Sigma}}^{-1}\left(\boldsymbol{\Sigma}^{-1}\boldsymbol{\mu} + \boldsymbol{\Sigma}_{\eta}^{-1}\boldsymbol{\eta} + \frac{\boldsymbol{\tau}_1}{\sigma_f^2}q_f + \frac{\boldsymbol{\tau}_2}{\sigma_{etf}^2}q_{etf}\right), \\ \hat{\boldsymbol{\Sigma}} &= \left(\boldsymbol{\Sigma}^{-1} + \boldsymbol{\Sigma}_{\eta}^{-1} + \frac{\boldsymbol{\tau}_1\boldsymbol{\tau}_1^T}{\sigma_f^2} + \frac{\boldsymbol{\tau}_2\boldsymbol{\tau}_2^T}{\sigma_{etf}^2}\right)^{-1}.\end{aligned}$$

After forming their posterior beliefs, investors select the portfolio that maximizes their utility. The sequence of events is shown in Figure 3.3.

3.3.3 Investors' Portfolio Choices

The stock prices in the first period are denoted by $\boldsymbol{P}$. Let P_f denote the futures price and P_{etf} denote the ETF price in the first period. These are the prices that need to be determined in equilibrium. We assume investors' utility is exponential in terminal wealth W_2,

$$U = E\left(\exp\left(-\gamma W_2\right)\right),$$

where γ is the coefficient of risk aversion. We assume the risk aversion for stock investors, futures providers and ETF providers to be respectively γ_s, γ_1 and γ_2. Let U_s, U_1, U_2 denote the utility functions of these investors. As the payoffs of stocks and derivatives are normally distributed, investors' wealth as a linear combination of assets is also normally distributed.

Accordingly, all investors' utility is equivalent to a mean-variance utility:

$$U = E[W_2] - \frac{\gamma}{2} Var[W_2] \; .$$

To determine equilibrium stock, futures and ETF prices, we solve the portfolio choice problem for each type of investor and then require markets to clear.

Let $\boldsymbol{X}_s$ denote stock investors' stock holdings and B_1^s their riskless asset holdings at $t = 1$. Stock investors only trade stocks and bonds at $t = 1$. Their wealth must satisfy

$$W_1 \;=\; \boldsymbol{P}^T \boldsymbol{X}_s + B_1^s = B_0^s + \frac{\boldsymbol{P}^T \boldsymbol{Q}}{N_s} - \frac{1}{2}(\boldsymbol{X}_s - \frac{\boldsymbol{Q}}{N_s})^T \boldsymbol{C}(\boldsymbol{X}_s - \frac{\boldsymbol{Q}}{N_s}), \tag{3.1}$$

$$W_2 \;=\; \boldsymbol{f}^T \boldsymbol{X}_s + B_0^s - \boldsymbol{P}^T \boldsymbol{X}_s + \frac{\boldsymbol{P}^T \boldsymbol{Q}}{N_s} - \frac{1}{2}(\boldsymbol{X}_s - \frac{\boldsymbol{Q}}{N_s})^T \boldsymbol{C}(\boldsymbol{X}_s - \frac{\boldsymbol{Q}}{N_s}). \tag{3.2}$$

In the first period, in equation (3.1), the value of the portfolio (stocks and bonds) equals the initial endowment minus trading costs, while in equation (3.2) in period 2, the value of the portfolio is the sum of the stock price (which equals its payoff) and the value of bonds. Because the gross interest rate is assumed to be 1, the value of the bonds in the second period is equal to their value in the first period and can be easily solved in equation (3.1) in terms of $\boldsymbol{X}_s$ and $\boldsymbol{P}$. Inserting W_2 into the utility function and taking expectations using posterior beliefs yields the following objective function for stock investors:

$$U_s = \hat{\boldsymbol{\mu}}^T \boldsymbol{X}_s + B_0^s - \boldsymbol{P}^T (\boldsymbol{X}_s - \frac{\boldsymbol{Q}}{N_s}) - \frac{1}{2}(\boldsymbol{X}_s - \frac{\boldsymbol{Q}}{N_s})^T \boldsymbol{C}(\boldsymbol{X}_s - \frac{\boldsymbol{Q}}{N_s}) - \frac{\gamma_s}{2} \boldsymbol{X}_s^T \hat{\boldsymbol{\Sigma}} \boldsymbol{X}_s. \tag{3.3}$$

The first-order condition (FOC) with respect to $\boldsymbol{X}_s$ is

$$\frac{\partial U_s}{\partial \boldsymbol{X}_s} = \hat{\boldsymbol{\mu}} - \boldsymbol{P} - \boldsymbol{C}(\boldsymbol{X}_s - \frac{\boldsymbol{Q}}{N_s}) - \gamma_s \hat{\boldsymbol{\Sigma}} \boldsymbol{X}_s = 0. \tag{3.4}$$

Solving, the optimal investment in stocks is

$$\boldsymbol{X}_s^* = (\boldsymbol{C} + \gamma_s \hat{\boldsymbol{\Sigma}})^{-1} (\hat{\boldsymbol{\mu}} - \boldsymbol{P} + \frac{1}{N_s} \boldsymbol{C}\boldsymbol{Q}) \; . \tag{3.5}$$

Stock investors take more aggressive positions in stocks when their expected return is high, their price is low and their variance is low. In the presence of non-zero transaction costs, the initial endowment $\boldsymbol{Q}$ also plays a role in determining optimal stock holdings. The higher the transaction costs are, the closer the optimal position is to the initial holdings.

Consider now futures arbitrageurs' portfolio choice. Let $\boldsymbol{X}_1$ denote their stock holdings, y their futures holdings and B_1^1 their riskless asset holdings in the first period. We further assume that $\boldsymbol{X}_1 = -\boldsymbol{b}x$, i.e., futures providers hedge their position by trading stocks in proportion to the index weight.[5] Because futures represent an unfunded position, there is no cash flow at $t = 1$. In period 2, the payoff per contract is $\boldsymbol{f}^T\boldsymbol{b} - P_f$. Hence, futures traders' wealth satisfies

$$
\begin{aligned}
W_1 &= \boldsymbol{P}^T\boldsymbol{b}x + 0y + B_1^1 = B_0^1 - \frac{1}{2}\boldsymbol{b}^T\boldsymbol{C}\boldsymbol{b}x^2 - \frac{1}{2}c_f y^2, && (3.6)\\
W_2 &= \boldsymbol{f}^T\boldsymbol{b}x + (\boldsymbol{f}^T\boldsymbol{b} - P_f)y + B_0^1 - \frac{1}{2}\boldsymbol{b}^T\boldsymbol{C}\boldsymbol{b}x^2 - \boldsymbol{P}^T\boldsymbol{b}x - \frac{1}{2}c_f y^2\,. && (3.7)
\end{aligned}
$$

In the first period, in equation (3.6), the value of the portfolio (stocks, bonds and futures) equals the initial endowment minus trading costs. In equation (3.7), the wealth in period 2 is the sum of stocks' payoff, the futures payoff $(\boldsymbol{f}^T\boldsymbol{b} - P_f)$ and the value of bonds. Using posterior beliefs, futures providers try to maximize

$$
U_1 = (\hat{\boldsymbol{\mu}} - \boldsymbol{P})^T\boldsymbol{b}x + (\hat{\boldsymbol{\mu}}^T\boldsymbol{b} - P_f)y - \frac{1}{2}\boldsymbol{b}^T\boldsymbol{C}\boldsymbol{b}x^2 - \frac{\gamma_1}{2}\boldsymbol{b}^T\hat{\boldsymbol{\Sigma}}\boldsymbol{b}(x+y)^2 - \frac{1}{2}c_f y^2. \tag{3.8}
$$

The FOCs with respect to x and y are

$$
\begin{aligned}
\frac{\partial U_1}{\partial x} &= (\hat{\boldsymbol{\mu}} - \boldsymbol{P})^T\boldsymbol{b} - \boldsymbol{b}^T\boldsymbol{C}\boldsymbol{b}x - \gamma_1\boldsymbol{b}^T\hat{\boldsymbol{\Sigma}}\boldsymbol{b}(x+y)\,, && (3.9)\\
\frac{\partial U_1}{\partial y} &= (\hat{\boldsymbol{\mu}}^T\boldsymbol{b} - P_f) - \gamma_1\boldsymbol{b}^T\hat{\boldsymbol{\Sigma}}\boldsymbol{b}(x+y) - c_f y\,. && (3.10)
\end{aligned}
$$

Solving and using the fact that $\boldsymbol{X}_1 = \boldsymbol{b}x$, the optimal positions in stocks and futures are

$$
\boldsymbol{X}_1^* = \frac{c_f\hat{\boldsymbol{\mu}}^T\boldsymbol{b} + \gamma_1\boldsymbol{b}^T\hat{\boldsymbol{\Sigma}}\boldsymbol{b}P_f - (\gamma_1\boldsymbol{b}^T\hat{\boldsymbol{\Sigma}}\boldsymbol{b} + c_f)\boldsymbol{P}^T\boldsymbol{b}}{c_f\boldsymbol{b}^T\boldsymbol{C}\boldsymbol{b} + (\boldsymbol{b}^T\boldsymbol{C}\boldsymbol{b} + c_f)\gamma_1\boldsymbol{b}^T\hat{\boldsymbol{\Sigma}}\boldsymbol{b}}\boldsymbol{b}, \tag{3.11}
$$

$$
y^* = \frac{-(\gamma_1\boldsymbol{b}^T\hat{\boldsymbol{\Sigma}}\boldsymbol{b} + \boldsymbol{b}^T\boldsymbol{C}\boldsymbol{b})P_f + \boldsymbol{b}^T\boldsymbol{C}\boldsymbol{b}\hat{\boldsymbol{\mu}}^T\boldsymbol{b} + \gamma_1\boldsymbol{b}^T\hat{\boldsymbol{\Sigma}}\boldsymbol{b}\boldsymbol{P}^T\boldsymbol{b}}{c_f\boldsymbol{b}^T\boldsymbol{C}\boldsymbol{b} + (\boldsymbol{b}^T\boldsymbol{C}\boldsymbol{b} + c_f)\gamma_1\boldsymbol{b}^T\hat{\boldsymbol{\Sigma}}\boldsymbol{b}}. \tag{3.12}
$$

The optimal positions in index futures and stocks are driven not only by their expected payoff $\hat{\boldsymbol{\mu}}$ but also by the price difference between futures and their underlying basket. When stocks' expected payoff is high, investors tend to buy more futures and stocks. When stocks are overpriced compared to futures, investors take more aggressive long positions in futures and short stocks more aggressively, and vice versa. Moreover, transaction costs also play a role in determining the size of positions. Higher transaction costs for stocks lead to less aggressive positions being taken in stocks, while higher futures' trading costs result in less aggressive holding in futures.

Turning to ETF traders, let $\boldsymbol{X}_2$ denote their stock holdings, z their ETF holdings and B_1^2 their riskless asset holdings in the first period. Because the ETF physically replicates the index, we

[5]This assumption simplifies the solution to the optimal choice problem. It is not a critical assumption.

have $\boldsymbol{X}_2 = -\boldsymbol{b}z$. For example, when ETF traders sell z units of the ETF to index investors (with $z < 0$), they must buy $-z\boldsymbol{b}$ stocks to build the ETF. This assumption is in line with current market practice in which most index ETFs physically replicate the index. Unlike writers of futures and other derivatives who can choose to hedge their positions only partially by trading stocks, the full replication characteristic of ETFs implies a strict relationship between ETFs and stock holdings for ETF traders. Accordingly, the wealth of ETF traders must satisfy

$$W_1 = -\boldsymbol{P}^T\boldsymbol{b}z + P_{etf}z + B_1^2 = B_0^2 - \frac{1}{2}\boldsymbol{b}^T\boldsymbol{C}\boldsymbol{b}z^2 - \frac{1}{2}c_{eft}z^2, \tag{3.13}$$

$$W_2 = B_0^2 - \frac{1}{2}\boldsymbol{b}^T\boldsymbol{C}\boldsymbol{b}z^2 - \frac{1}{2}c_{eft}z^2 + \boldsymbol{P}^T\boldsymbol{b}z - P_{etf}z. \tag{3.14}$$

In the first period, the budget constraint (3.13) requires portfolio holdings to be equal to the initial endowment minus trading costs. In the second period, the payoffs of the stocks and the ETF are identical. Because ETF providers take offsetting positions in the ETF and the stocks, these two terms cancel out in equation (3.14). The wealth in the second period equals the value of the bond holdings. ETF providers maximize

$$U_2 = -(P_{etf} - \boldsymbol{P}^T\boldsymbol{b})z - \frac{1}{2}\boldsymbol{b}^T\boldsymbol{C}\boldsymbol{b}z^2 - \frac{1}{2}c_{eft}z^2. \tag{3.15}$$

Because stocks' payoffs offset the payoff of the ETF, posterior beliefs do not enter the objective function.

The FOC with respect to z is

$$\frac{\partial U_2}{\partial z} = -P_{etf} - \boldsymbol{P}^T\boldsymbol{b} - (\boldsymbol{b}^T\boldsymbol{C}\boldsymbol{b} + c_{eft})z\,. \tag{3.16}$$

The optimal positions in stocks and the ETF are

$$\boldsymbol{X}_2^* = -\frac{\boldsymbol{P}^T\boldsymbol{b} - P_{etf}}{\boldsymbol{b}^T\boldsymbol{C}\boldsymbol{b} + c_{etf}}\boldsymbol{b}, \tag{3.17}$$

$$z^* = \frac{\boldsymbol{P}^T\boldsymbol{b} - P_{etf}}{\boldsymbol{b}^T\boldsymbol{C}\boldsymbol{b} + c_{etf}}. \tag{3.18}$$

Optimal holdings in stocks are the optimal ETF holding multiplied by the index weight $\boldsymbol{b}$. Both ETF and stock holdings are determined by the price difference between these two assets. When stock prices are high, ETF providers take more aggressive short positions in stocks and a long position in the ETF and vice versa. Higher trading costs for stocks or the ETF decrease the position size in both stocks and the ETF.

3.3.4 Equilibrium Prices

Having solved the portfolio choice problems for all three types of investors in the previous subsection, we can derive the equilibrium prices by requiring that the stock, futures and ETF markets clear, i.e.,

$$N_s \boldsymbol{X}_{\boldsymbol{s}}^* + N_1 \boldsymbol{X}_1^* + N_2 \boldsymbol{X}_2^* = \boldsymbol{Q}, \tag{3.19}$$
$$N_1 y^* + q_f = 0, \tag{3.20}$$
$$N_2 z^* + q_{etf} = 0. \tag{3.21}$$

Inserting the optimal portfolio choice for all investors presented in equations (3.5), (3.11), (3.12), (3.17) and (3.18) in the market clearing conditions, we show in Appendix 3.A that equilibrium stock prices are given by:

$$\boldsymbol{P} = \hat{\boldsymbol{\mu}} + \alpha \boldsymbol{D}^{-1} \boldsymbol{b} q_f + \boldsymbol{D}^{-1} \boldsymbol{b} q_{etf} + \boldsymbol{D}^{-1} \left((\boldsymbol{C} + \gamma_s \hat{\boldsymbol{\Sigma}})^{-1} \boldsymbol{C} - \boldsymbol{I} \right) \boldsymbol{Q}, \tag{3.22}$$

where $\boldsymbol{D} = N_s(\boldsymbol{C} + \gamma_s \hat{\boldsymbol{\Sigma}})^{-1} + \lambda \boldsymbol{b}\boldsymbol{b}^T$, $\alpha = \frac{\gamma_1 \boldsymbol{b}^T \hat{\boldsymbol{\Sigma}} \boldsymbol{b}}{\gamma_1 \boldsymbol{b}^T \hat{\boldsymbol{\Sigma}} \boldsymbol{b} + \boldsymbol{b}^T \boldsymbol{C} \boldsymbol{b}}$ and $\lambda = N_1 \left(c_f + \frac{\gamma_1 \boldsymbol{b}^T \hat{\boldsymbol{\Sigma}} \boldsymbol{b} \boldsymbol{b}^T \boldsymbol{C} \boldsymbol{b}}{\gamma_1 \boldsymbol{b}^T \hat{\boldsymbol{\Sigma}} \boldsymbol{b} + \boldsymbol{b}^T \boldsymbol{C} \boldsymbol{b}} \right) / a$.

Equilibrium stock prices increase one for one with their expected payoff $\hat{\boldsymbol{\mu}}$ at $t = 1$. In addition to expected stock returns, stock prices are affected by q_f and q_{etf}. Comparing the coefficients for q_f and q_{etf}, we can see immediately that the impact of ETF and futures demand shocks is the same up to scaling. Moreover, because $0 \leq \alpha \leq 1$, futures demand shocks q_f have a smaller impact on stock prices than ETF shocks q_f. When there are transaction costs, futures providers do not hedge their futures position completely with stocks, while ETF providers are required to fully hedge their ETF holdings because of the full physical replication property. When there are no transaction costs for stocks ($\boldsymbol{C} = \boldsymbol{0}$), futures traders also hedge their futures positions perfectly using stocks. In this case, $\alpha = 1$ and demand shocks from the futures market and ETFs have the same price impact on stocks. However, whenever $\boldsymbol{C} \neq \boldsymbol{0}$, $\alpha < 1$ and ETF demand shocks have a larger price impact than futures demand shocks.

We now investigate whether demand shocks affect both index and non-index stock prices. We distinguish two cases. In the special case when stocks have independent liquidity ($\boldsymbol{C}$ is a diagonal matrix) and independent posterior beliefs ($\hat{\boldsymbol{\Sigma}}$ is a diagonal matrix), demand shocks only impact index stocks. Non-index stock prices are the same as in the case without demand shocks. When the demand shock is positive, the prices of all index stocks increase and vice versa. This result is proved in Appendix 3.B. However, if the liquidity or posterior beliefs of non-index stocks are correlated with those of index stocks, demand shocks may also affect the prices of non-index stocks. For example, for a stock i ($> k$) that is not in the index, $b_i = 0$. However, the ith element of the vector $\boldsymbol{D}^{-1}\boldsymbol{b}$ could be non-zero; this could happen when stock i has a non-zero cross term with one or more index stocks in matrices $\boldsymbol{C}$, $\boldsymbol{\Sigma}$ or $\boldsymbol{\Sigma}_\eta$. In such a case, although stock i is not in the index, its price is also influenced by demand shocks that hit the index market. It is difficult to make a general statement about which non-index stocks are affected. However, as long as the submatrix formed by taking the first k columns and last $N-k$ rows of matrix $\boldsymbol{D}^{-1}$ have non-zero elements, $\boldsymbol{D}^{-1}\boldsymbol{b}$ will have a non-zero element for non-index stocks, and non-index stocks will also be affected by demand shocks in the ETF and futures markets.

We now analyze the impact of demand shocks on risk premia. Risk premia are defined as expected payoffs minus the prices growing at the risk-free rate. Plugging the equilibrium price into equation (3.22) and using the fact that $r = 1$, we obtain the following expression for risk premia:

$$\mathbb{E}[\boldsymbol{f} - \boldsymbol{P}r] \quad = \quad -\alpha \boldsymbol{D}^{-1}\boldsymbol{b}\boldsymbol{\tau}_{\mathbf{1}}^T \hat{\boldsymbol{\mu}} - \boldsymbol{D}^{-1}\boldsymbol{b}\boldsymbol{\tau}_{\mathbf{2}}^T \hat{\boldsymbol{\mu}} - \boldsymbol{D}^{-1}\left((\boldsymbol{C} + \gamma_s \hat{\boldsymbol{\Sigma}})^{-1}\boldsymbol{C} - \boldsymbol{I}\right)\boldsymbol{Q} \,. \tag{3.23}$$

In the special case where liquidity and posterior beliefs are a diagonal matrix, index stocks offer lower premia than in the case without demand shocks, while risk premia on non-index stocks are unaffected. These premia respond differently because although stock investors require risk premia to hold index stocks, arbitrageurs do not require risk premia from index stocks when they trade stocks to reap arbitrage profits. Hence, risk premia offered by index stocks decrease. Because non-index stocks do not play a role in arbitrageurs' portfolio, both stock investors and arbitrageurs require risk premia to hold non-index stocks. Thus, risk premia for non-index stocks are unaffacted. However, when non-index stocks have correlated liquidity or fundamentals with index stocks, we cannot make a general statement regarding the direction of the risk premia. Risk premia of non-index stocks are likely to be affected by demand shocks as well.

3.3.5 Impact on Volatility and Correlations

When there are no demand shocks, equilibrium stock prices can be derived simply by setting $q_f = 0$ and $q_{etf} = 0$ in equation (3.22). Denote $\bar{\boldsymbol{P}}$ the equilibrium stock price and $\bar{\boldsymbol{\mu}}$ the posterior belief without demand shocks. Equilibrium stock prices without demand shocks are given by

$$\bar{\boldsymbol{P}} = \bar{\boldsymbol{\mu}} + \boldsymbol{D}^{-1}\left((\boldsymbol{C} + \gamma_s \hat{\boldsymbol{\Sigma}})^{-1}\boldsymbol{C} - \boldsymbol{I}\right)\boldsymbol{Q} \,. \tag{3.24}$$

As can be seen in equation (3.22), in the presence of demand shocks, there are more sources of uncertainty in equilibrium stock prices. However, the impact of these shocks on price volatility and correlation is not obvious. First considering volatility, stock price volatility without and with demand shocks is given by

$$V(\bar{\boldsymbol{P}}) = V(\bar{\boldsymbol{\mu}}), \tag{3.25}$$

$$\begin{aligned} V(\boldsymbol{P}) = & \underbrace{V(\hat{\boldsymbol{\mu}})}_{(A)} + \underbrace{\alpha^2\sigma_f^2\boldsymbol{D}^{-1}\boldsymbol{b}\boldsymbol{b}^T\boldsymbol{D}^{-1} + \sigma_{etf}^2\boldsymbol{D}^{-1}\boldsymbol{b}\boldsymbol{b}^T\boldsymbol{D}^{-1}}_{(B)} \\ & + \underbrace{\alpha COV(\hat{\boldsymbol{\mu}}, \boldsymbol{D}^{-1}\boldsymbol{b}q_f) + COV(\hat{\boldsymbol{\mu}}, \boldsymbol{D}^{-1}\boldsymbol{b}q_{etf})}_{(C)} . \end{aligned} \tag{3.26}$$

Price volatility with demand shocks has three components. Part (A) represents the uncertainty from the posterior belief. Part (B) is the uncertainty due to the variance of demand shocks. Part (C) is the covariance between beliefs and demand shocks. It is not immediately obvious which of the two volatility expressions in equations (3.25) and (3.26) is larger. We can, however, compare these two quantities in the special case when demand shocks contain pure noise, i.e., when $\boldsymbol{\tau}_1$ and $\boldsymbol{\tau}_2$ are zero. Because demand shocks do not convey information about future payoffs, only the public signal is useful for updating the belief. Accordingly, stock investors and arbitrageurs update their beliefs using

$$\hat{\boldsymbol{\mu}} = \hat{\boldsymbol{\Sigma}}^{-1}\left(\boldsymbol{\Sigma}^{-1}\boldsymbol{\mu} + \boldsymbol{\Sigma}_\eta^{-1}\boldsymbol{\eta}\right) .$$

It is obvious that in this special case, the posterior belief $\hat{\boldsymbol{\mu}}$ is equal to the posterior belief $\bar{\boldsymbol{\mu}}$ in the case where there are no demand shocks . Because demand shocks play no role in determining poster beliefs, COV$(\hat{\boldsymbol{\mu}}, \boldsymbol{D}^{-1}\boldsymbol{b}q_{etf}) = 0$ and COV$(\hat{\boldsymbol{\mu}}, \boldsymbol{D}^{-1}\boldsymbol{b}q_f) = 0$. Hence, we obtain the following:

$$V(\boldsymbol{P}) = V(\bar{\boldsymbol{P}}) + \alpha^2\sigma_f^2\boldsymbol{D}^{-1}\boldsymbol{b}\boldsymbol{b}^T\boldsymbol{D}^{-1} + \sigma_{etf}^2\boldsymbol{D}^{-1}\boldsymbol{b}\boldsymbol{b}^T\boldsymbol{D}^{-1} . \tag{3.27}$$

Because $\alpha^2\sigma_f^2\boldsymbol{D}^{-1}\boldsymbol{b}\boldsymbol{b}^T\boldsymbol{D}^{-1}$ and $\sigma_{etf}^2\boldsymbol{D}^{-1}\boldsymbol{b}\boldsymbol{b}^T\boldsymbol{D}^{-1}$ are positive definite, the diagonal elements of $V(\boldsymbol{P})$ are larger than those of $V(\bar{\boldsymbol{P}})$. Demand shocks simply add more noise to the price when they contain no information, i.e., stock prices with demand shocks have higher volatility than prices without demand shocks.

However, if demand shocks contain some information about the payoff at $t = 2$, we cannot make a definite statement about the direction of the volatility change because the additional precision from having more information may compensate for the noise that demand shocks add to stock prices. In other words, part (A) of the variance might decrease in the presence of demand shocks, although part (B) of the variance is higher compared to the case without demand shocks.

Turning now to correlations, without demand shocks on index products, stock prices are only correlated due to correlations in posterior beliefs about their payoffs. To be more precise,

$$COV(\bar{P}_i, \bar{P}_j) = COV(\bar{\mu}_i, \bar{\mu}_j) \ . \tag{3.28}$$

When the posterior expectations for stocks i and j are correlated, their prices are correlated. Because posterior beliefs combine both ex-ante beliefs and the public signal, and $\boldsymbol{\mu}$ is constant, the covariance between $\bar{\mu}_i$ and $\bar{\mu}_j$ is determined by the public signal.

In the presence of demand shocks, correlations are also driven by stocks' exposures to demand shocks. However, it is not clear whether demand shocks cause correlations to increase or decrease because demand shocks not only enter as additional factors in equilibrium prices but also affect posterior beliefs. We can only make unambiguous predictions in the special case when $\boldsymbol{\Sigma}^{-1}$ and $\boldsymbol{\Sigma}_\eta^{-1}$ are diagonal matrices and the demand shocks are composed of pure noise. Formally, in this case, the prices of two stocks i and j can be written as

$$P_i = \mu_i + a_2\eta_i + a_3 q_f + a_4 q_{etf} + \boldsymbol{a}_5\boldsymbol{Q} \ , \tag{3.29}$$

$$P_j = \mu_j + b_2\eta_j + b_3 q_f + b_4 q_{etf} + \boldsymbol{b}_5\boldsymbol{Q} \ . \tag{3.30}$$

Because η_i and η_j are not correlated, correlation only arises from the common exposure to q_f and q_{etf}. If both stocks have positive (negative) loadings on demand shocks, then both stocks have a correlation that exceeds the correlation without demand shocks (which is zero because the fundamentals are uncorrelated). Otherwise, they will be negatively correlated.

3.4 Empirical Tests

The model in Section 3.3 demonstrates that demand shocks to index ETFs and futures affect stocks' equilibrium prices and their correlations. Hence, demand shocks should also play a role in the correlation structure of stock returns. In this section, we investigate the relationship between futures and ETF demand shocks and stock return correlations. Because broad-based futures and ETFs are likely to affect stock return correlations as a whole, we focus on their impact at an aggregate level rather than on specific pairwise correlations.

3.4.1 Data

We obtain the composition of the S&P 500 index from the Center for Research in Security Prices (CRSP). Daily S&P 500 constituents' stock prices, shares outstanding and trading volume are also from CRSP. For non-S&P 500 stocks, we use those stocks for which CRSP provides beta estimates and that have a market capitalization of at least 100 million dollars and a share price of at least 5 dollars. Our sample comprises 5880 stocks. We also obtain price, trading volume and shares outstanding from CRSP for the three largest ETFs that track S&P

500 stocks, namely SPY, IVV and VOO. For futures, we collect daily futures prices and trading volume from Bloomberg. Because futures were first introduced on April 21, 1982, we choose the period from April 1982 to December 2012 for the analysis considering all futures. For the analysis based on legacy futures, we consider the period from April 1982 to August 1997, as legacy futures were gradually replaced by E-mini futures after 1997. For the analysis based on E-mini futures, we investigate the period from September 1997 to December 2012. Finally, for ETFs, we analyze the period from January 1993, when SPY was introduced, to December 2012.

Although sector ETFs constructed to reflect broad industries are available, our analysis only considers ETFs on the S&P 500 index because sector ETFs are roughly only half of the size of S&P 500 index ETFs, so their impact on aggregate correlations should be much smaller than that of S&P 500 ETFs.

3.4.2 Correlation among Stocks

Because our focus is to find evidence of the impact of index demand shocks at an aggregate level and not to investigate whether this impact differs across stocks, we follow Pollet and Wilson (2010) and study a measure of aggregate stock return correlation. To be more precise, for each month t, we compute the S&P 500 average (value weighted) correlation using daily stock returns as follows:

$$\rho_t = \Sigma_{i=1}^{500}\Sigma_{j=1}^{500} w_{it} w_{jt} \rho_{ijt}\,, \tag{3.31}$$

where ρ_{ijt} is the Pearson stock return correlation between stocks i and j, and w_{it} and w_{jt} are their weights in month t. This average correlation ρ_t lies between -1 and 1 and will be 1 when all pairwise stock returns are perfectly correlated.

We choose to compute correlations at a monthly frequency to have a reasonable number of observations, given that we only have 15 years of data for E-mini futures; however, the results presented below are robust if we use quarterly correlations. Unlike rolling windows, using non-overlapping windows avoids the problem of having a few observations that affect m consecutive periods, with m being the window span. For example, if we compute 1-year rolling correlations, the correlation jumps on October 19, 1987, due to the market crash and remains high for an entire year, with a peak almost a year later on September 29, 1988. The rolling correlation only returns to its pre-cash level two years after the crash. By contrast, the monthly correlation returns to its pre-crash level a few months after the crash. We follow French, Schwert, and Stambaugh (1987) and calculate the monthly S&P 500 volatility using daily returns during the month. Letting N_t denote the number of trading days, the sample

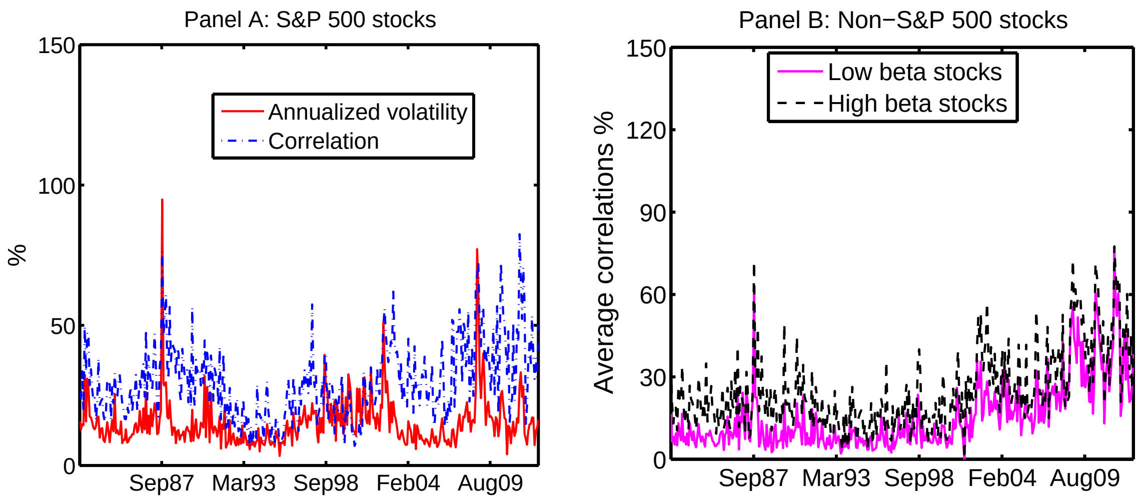

FIGURE 3.4: *Average correlation and realized volatility over time.*
Panel A shows the monthly average correlation of S&P 500 stocks and S&P 500 realized volatility over time. Panel B compares the average correlation of high-beta stocks and low-beta stocks over time.

volatility of the S&P 500 index for a given month t is

$$\text{SPVol}_t \quad = \quad \left(\Sigma_{i=1}^{N_t} r_{it}^2 + 2\Sigma_{i=1}^{N_t-1} r_{it} r_{i+1,t}\right)^{\frac{1}{2}}, \tag{3.32}$$

where r_{it} is the S&P 500 daily return on day i of month t. The cross term is included to control for return auto-correlation.

Figure 3.4 reports monthly value weighted average correlations among S&P 500 stocks and monthly S&P 500 realized volatility during the sample period. As can be seen in Panel A, the average correlation usually lies between 10 and 60 percent. Towards the end of the sample period, the average correlation was at historically high levels, with three recent bursts of correlation, namely the Lehman default, the Flash Crash and the downgrade of United States sovereign debt by Standard and Poor's. These high correlations were matched in intensity only during the 1987 market crash. Figure 3.4 also shows that correlations move mostly in the same direction as volatility over time. For example, during the market crash in 1987, both volatility and correlation spike. High volatility and correlation can also be observed during the recent financial crisis, when volatility reaches 77% and the average correlation jumps to 73%. However, the average level of correlation had been gradually increasing over the previous ten years; it has remained high after the crisis, despite the fact that volatility has returned to a low level. This result shows that even though correlation levels ebb and flow with volatility cycles, correlations can sometimes decouple from volatility.

As our model shows that correlations between non-index stocks are also affected by demand shocks to index derivatives, we also compute the average correlation of non-S&P 500 stocks.

However, non-index stocks are not directly included in the index and also have correlated fundamentals with index stocks. Hence, non-index stocks would play a hedging role in investors' portfolios. The price of these stocks would also change because they hedge investors' positions in the affected securities.

Due to the large number of non-S&P 500 stocks, we further divide these stocks into two equally sized groups according to their market beta. Stocks with high market beta are grouped in the high-beta category, while stocks with lower market beta are included in the low-beta group. Similar to the computation of the S&P 500 correlation, we obtain the average correlations within the high and low beta groups using equation (3.31), where the weight w_{it} is given by the market capitalization of stock i divided by the total market capitalization of its group.

Panel B of Figure 3.4 indicates that the evolution of correlations among non-index stocks over time is similar to that of S&P 500 stocks. Furthermore, the average correlations of the high and low beta groups move together over time. Stocks with higher beta have higher average correlations with each other than stocks with lower beta.

3.4.3 Construction of the Dependent and Explanatory Variables

We use the monthly average correlation ρ_t as the dependent variable. Because this correlation measure, computed in Section 3.4.2, lies between -1 and 1, we perform a Fisher transformation to obtain a modified correlation measure that takes values on the entire real line,

$$\bar{\rho}_t = \frac{1}{2}\ln\left(\frac{1+\rho_t}{1-\rho_t}\right) . \tag{3.33}$$

This modified correlation $\bar{\rho}_t$ is a strictly monotone transform of the actual correlation and is employed as the dependent variable in our empirical analysis. The modified correlation is close to the average Pearson correlation ρ_t when the latter lies between -0.5 and 0.5. For larger values, the Fisher transform might result in modified correlations exceeding 1.

Because the dependent variable is computed at a monthly frequency, all explanatory variables are also constructed for each month. The results in Section 3.3 indicate that correlations in stock prices are related to demand shocks to index products. However, these demand shocks are not directly observable and are hard to disentangle from other sources of trading activity. As a proxy for demand shocks, we use the ratio of the dollar trading volume of index futures or ETFs to the aggregate dollar trading volume of S&P 500 stocks during each month:

$$\text{Trading Ratio} = \frac{\text{ETF (futures) dollar trading volume}}{\text{S\&P 500 dollar trading volume}} .$$

This ratio directly compares the volume in both markets and has several advantages. First, using dollar trading volume automatically controls for the size of the contract. Contract sizes for E-mini futures, legacy futures and ETFs are currently 50, 250 and 1/10, respectively. Second, dividing ETF or futures volume by the dollar trading volume of the S&P 500 controls for other potential drivers of comovement. For example, Barberis, Shleifer, and Wurgler (2005) argue that investors trade stocks based on the 'category view' or 'habit view', i.e., they buy or sell all stocks that they consider to be in the same category. If stocks are correlated for that reason, the aggregate stock trading volume of the stocks will be high and our trading ratio will tend to be lower. Third, the trading ratio is likely to increase in the presence of a demand shock that hits only index products. When a demand shock hits both the stock market and index products, the ratio does not necessarily increase. Hence, the ratio constitutes a good proxy for demand shocks that hit only the index derivatives market.[6]

The literature explaining the time variation of average stock market correlations is sparse and has only identified a few factors that might explain stock return correlations. We incorporate these factors as well as several economic variables that are commonly used in explaining stock returns and volatilities as controls in our analysis. We now describe the stock market, bond market and macroeconomic variables that we use.

From the stock market, we include the realized volatility SPVol_t and the current month return for the S&P 500 because there is evidence that correlations increase in bear markets and volatile markets (see, for example, Longin and Solnik (1995) and Ang and Chen (2002)).

From the bond market, we include the 3 month treasury bill rate and the default spread. During our sample period, the interest rate is not stationary, and thus the difference in the interest rate is taken as the independent variable. As the default spread, we use the yield difference between AAA bonds and BAA bonds from CRSP; their difference has historically been a good indicator of whether the economy was in recession or expansion.

We also include several macroeconomic variables such as industrial production growth, policy uncertainty, inflation and sentiment. We include these quantities because correlations have been found to depend on the phase of the business cycle and are high during recessions (Ledoit, Santa-Clara, and Wolf (2003) and Erb, Harvey, and Viskanta (1994)). We collect data on the growth rate of US industrial production to control for growth in the economy. Compared with GDP growth, industrial production has the advantage of being available monthly. Because the impact of economic growth might be reflected slowly in financial markets, we use the following variable in the regression:

$$\Delta\text{IndustrialProduction}_t = \log\left(\Sigma_{i=t-11}^{t}\text{IndustrialProduction}_i\right) - \log\left(\Sigma_{j=t-23}^{t-11}\text{IndustrialProduction}_j\right).$$

Basically, we take the log difference between industrial production during the last 12 months and that during the prior 12 months. The impact of policy uncertainty on different economic variables such as growth, inflation and investment has been studied extensively in the literature (see Bloom (2009), Bachmann, Elstner, and Sims (2010), and Jones and Olson (2013)

[6] For ETFs, an alternative way to compare trading activity for the index ETF market and the underlying market is to compute turnover. Using the turnover ratio, we find results similar to those reported below; therefore, they are omitted for brevity.

among others). Because increased policy uncertainty negatively affects the stock market, it might increase stock return correlations. We obtain the policy uncertainty index from the website by Baker, Bloom and Davis. Inflation has an impact on stock returns. The previous literature finds a negative relationship between stock returns and inflation in the short term, but this relationship becomes positive in the long term (see Jaffe and Mandelker (1976) and Boudoukh and Richardson (1993) for an example). Regardless of the sign of this impact, inflation is likely to have a similar impact on all stocks and thus to increase stock return correlations. We take the Consumer Price Index (CPI) from the US Bureau of Labor Statistics and use the inflation rate as the independent variable. The inflation rate inflation_t is computed as the logarithm of the ratio of CPI values at t and $t-1$. We also include market sentiment, which is obtained from Baker and Wurgler's website. Their sentiment index is based on the common variation in six underlying proxies for sentiment. Sentiment is found to have an impact on the cross-section of stock returns (see Baker and Wurgler (2006), for example). For example, when sentiment is low, subsequent returns are higher for newly listed, more volatile and unprofitable stocks. Although sentiment's impact on stock returns varies, we still include it as a control variable.

The summary statistics of these variables are reported in Table 3.1 and their correlations in Table 3.2. As can be seen in Table 3.1, average stock return correlations are highest for index stocks and lowest for low-beta non-index stocks. Transformed correlations are on average close to Pearson correlations, but with much higher maximum values and skewness.

[Table 3.1 and Table 3.2 about here]

As can been seen from the first column of Table 3.2, the average stock return correlation is highly correlated with volatility, with a correlation of 62 percent. Its correlation with the other variables is generally much lower.

3.4.4 Hypotheses and Regression Specification

The goal of our empirical analysis is to assess whether index demand shocks affect stock return correlations. As demonstrated in Section 3.3.5, these demand shocks should affect both index and non-index stocks. Accordingly, in our empirical analysis, we test the following three hypotheses:

Hypothesis 1. ETF and futures trading activity affects the correlations of S&P 500 stocks.

Hypothesis 2. ETF and futures trading activity affects the correlations of non-S&P 500 stocks.

Hypothesis 3. ETF trading activity has a stronger impact on correlations than futures trading activity.

Before estimating the regressions to test the hypotheses, it is necessary to test the stationarity of all the variables to avoid making spurious inferences. To do so, we employ the Augmented

Dicky-Fuller (ADF) test with a lag selected based on the Schwarz information criterion. The values of the ADF test are reported in Table 3.1. The ADF test confirms the stationarity of average correlation $\bar{\rho}_t$ and realized monthly volatility SPVol_t.

However, the trading ratios for ETFs, futures and E-mini futures are found to be non-stationary.[7] We use three approaches to remove the non-stationarity of the trading ratios. The first is to subtract the average value during the previous year from the raw series. The resulting trading ratios are stationary. The second is to take the log of the growth in the trading ratio, i.e., $\log\left(\frac{\text{TradingRatio}_t}{\text{TradingRatio}_{t-1}}\right)$, to obtain a stationary time series. The third is to fit an ARIMA model with order selected using the Schwarz information criterion. The fitted ratio reflects activity that is forecastable but highly variable across months, while the residual of the time series represents the unexpected trading ratio, which we use as a regressor in our analysis. Although the results are the weakest when we use the trading ratios constructed using the second method, they are similar across methods. In the following, we therefore only report the results obtained when using the residual fitted using the ARIMA model.

As a benchmark, we first run a regression that only includes all control variables:

$$\begin{aligned}\bar{\rho}_t \quad = \quad & \beta_0 + \beta_1 \text{SPVol}_t + \beta_2 \text{CurrentReturn}_t + \beta_3 \Delta \text{3MonthTbill}_t + \beta_4 \text{CreditSpread}_t \\ & + \beta_5 \text{Termspread}_t + \beta_6 \Delta \text{IndustrialProduction}_t + \beta_7 \text{Inflation}_t + \beta_8 \text{PolicyUncertainty}_t \\ & + \beta_9 \bar{\rho}_{t-1} + \varepsilon_t. \end{aligned} \tag{3.34}$$

The variance inflation factor is below 5, suggesting that there is no multicollinearity issue with this specification.

We then add our proxy for demand shocks as an explanatory variable, i.e., estimate the regression

$$\begin{aligned}\bar{\rho}_t \quad = \quad & \beta_0 + \beta_1 \Delta \text{IndexTrading}_t + \beta_2 \text{SPVol}_t + \beta_3 \text{CurrentReturn}_t + \beta_4 \Delta \text{3MonthTbill}_t \\ & + \beta_5 \text{CreditSpread}_t + \beta_6 \text{Termspread}_t + \beta_7 \Delta \text{IndustrialProduction}_t + \beta_8 \text{Inflation}_t \\ & + \beta_9 \text{PolicyUncertainty}_t + \beta_{10} \bar{\rho}_{t-1} + \varepsilon_t. \end{aligned} \tag{3.35}$$

The coefficient β_1 measures the impact of index trading activity on stocks' average correlations. A significant and positive β_1 is evidence that index trading activity affects these correlations.

[7]Using the Zivot and Andrews unit root test, which allows for a structural break in the time series, gives similar results.

3.4.5 Results for S&P 500 Stocks

The estimation results for regressions (3.34) and (3.35) when using the average correlation of S&P 500 stocks are reported in Table 3.3. We use four different trading ratios as explanatory variables: all futures (sum of legacy and E-mini futures), legacy futures, ETFs and E-mini futures. Because these index products were introduced at different points in time, we perform the estimation for different time periods as reported in the header row in the table.

[Table 3.3 about here]

The coefficient of realized volatility is significant and positive in all regressions. This result is consistent with the observation in Section 3.4.2 that the average correlation and volatility generally move in the same direction.[8]

Although their significance is lower than that of volatility, other variables also affect average correlations. Specifically, a negative current market return is associated with higher average correlation. This result is in line with the observation that stocks are more strongly correlated during market turmoil. The coefficient of the credit spread is significantly positive in one of the subperiods, reflecting the fact that a higher spread reflects weak macroeconomic conditions, which tend to be associated with higher correlations. The results in Table 3.3 also confirm the intuition that the inflation rate is high when stock return correlations are high because inflation affects most stocks in a similar way. Indeed, the coefficient for Inflation_t is significantly positive for two of the subperiods. As expected, policy uncertainty is significantly positively related to average correlations. Other controls such as the T-bill rate, sentiment and industrial production are insignificant.

Focusing now on the trading ratios, we see that all trading ratios are significantly and positively related to average correlations. The coefficient is the highest for ETFs at 3.748, which is consistent with the model's prediction that demand shocks to ETFs should have a stronger impact because of physical replication. The coefficient for E-mini futures is around 0.612, much higher than the coefficient for legacy futures, which is 0.103. This difference is probably due to the fact that legacy futures are floor-based, and there are delays in execution. Therefore, the arbitrage relationship between E-mini futures and the underlying stocks is more likely to hold for E-mini futures than for legacy futures. In other words, a demand shock to E-mini futures is more likely to lead to arbitrage activity between the E-mini futures and the stock market than a shock to the legacy futures market. In the last column of Table 3.3, both ETFs and futures are included in the regression. It turns out that trading activity in both markets has a significant positive impact on average stock correlations.

When comparing the R^2 of the benchmark regression and the combined regression for the period during which only legacy futures are traded, the increase in R^2 achieved by adding

[8] We also estimate the regression using VIX as a measure of index volatility and find similar results, which are not reported for brevity.

futures trading activity is only approximately 1.3 percent. But for the entire period and periods when E-mini futures and ETFs are traded, including futures and ETF trading activity can lead to a large increase in R^2. The increase in R^2 is around 4 percent for the entire sample, 9 percent for the E-mini sample, and 17 percent for the ETF sample. Overall, the results show that a large part of the time variation in index stock correlations can be explained by index trading activity. This result holds even when controlling for a wide set of variables that are known to affect correlations.

3.4.6 Results for Non-S&P 500 Stocks

We now investigate whether futures and ETF trading activity also explains the time variation in correlations of non-index stocks. To accomplish this investigation, we run the benchmark regression (3.34) as well as the combined regression (3.35) using the average correlation of non-S&P 500 stocks as the independent variable. We categorize non-S&P 500 stocks into two groups according to their beta, i.e., build a high beta group and a low beta group. The regression results are reported in Table 3.4 for high-beta stocks and Table 3.5 for low-beta stocks.

[Table 3.4 about here]

[Table 3.5 about here]

As expected, the coefficients of the control variables have the same sign as for index stocks. In general, a higher correlation is associated with higher volatility, a lower current month return, a higher credit spread, lower industrial production growth, higher inflation and more policy uncertainty. One exception is that the credit spread is negatively associated with the correlation of high and low beta stocks for the period when legacy futures are traded (see Table 3.4 and 3.5) because during 1982-1997, correlations spiked during the 1987 crash, while the credit spread remained low in October 1987. Black Friday had a substantial impact on the relation between the correlation and the credit spread. During other periods, the credit spread is more likely to be positively linked with correlation, echoing the positive significant coefficients observed for other periods.

Although non-S&P stocks are not included in the index, Tables 3.4 and 3.5 show that their correlations can in part be explained by trading activity in index products. Except for the time period 1982-1997, when only floor-based futures were traded, the coefficient of index trading activity is significantly positive. The t-statistics are especially high for E-mini futures as well as for ETFs. Moreover, as was the case for index stocks, ETFs have higher coefficients than E-mini and floor-based futures. The coefficient for ETFs is above 2, while it is around 0.5 for E-mini futures. Turning to R^2, for the entire period, the increase in R^2 achieved by adding futures trading activity as an explanatory variable is only approximately 1 percent.

Surprisingly, as the last row of Tables 3.4 and 3.5 suggest, incorporating E-mini futures and ETF trading ratios significantly increases R^2 by 6 to 9 percent. Summarizing, trading activity in index products such as ETFs and E-mini futures is also important in explaining the time variation of non-index stocks' correlations. These results might appear to be surprising but are in line with the predictions of the model presented in Section 3.3.

Although index trading activity affects the correlations of both index and non-index stocks, one would expect the magnitude of the impact to differ between groups. To examine this hypothesis, we pool all monthly correlations (for index stocks, high beta stocks and low beta stocks) together in one regression. We also introduce a dummy variable I_{index} that indicates whether the observation relates to index stocks or not along with an interaction term $\mathrm{I}_{index} \cdot \Delta\mathrm{IndexTrading}_i$. Formally, we estimate the following regression:

$$\begin{aligned} \bar{\rho}_i \quad = \quad & \beta_0 + \beta_1 \Delta\mathrm{IndexTrading}_i + \beta_2 \mathrm{I}_{index} + \beta_3 \mathrm{I}_{index} \cdot \Delta\mathrm{IndexTrading}_i + \beta_4 \mathrm{SPVol}_i \\ & + \beta_5 \mathrm{CurrentReturn}_i + \beta_6 \Delta 3\mathrm{MonthTbill}_i + \beta_7 \mathrm{CreditSpread}_i + \beta_8 \mathrm{Termspread}_i \\ & + \beta_9 \mathrm{Inflation}_i + \beta_{10} \Delta\mathrm{IndustrialProduction}_i + \beta_{11} \mathrm{PolicyUncertainty}_i + \varepsilon_i. \quad (3.36) \end{aligned}$$

Coefficient β_3 measures the differential sensitivity of S&P stock correlations to index trading activity compared with non-index stocks. The results for different subperiods are presented in Table 3.6.

[Table 3.6 about here]

As can be seen in Table 3.6, index trading activity has a significant impact on correlations. Moreover, the coefficient for the index dummy is significantly positive, indicating that index stock correlations are significantly higher than non-index stock correlations. This result is in line with Figure 3.4, where the correlations of index stocks are higher than those of non-index stocks. More importantly, the interaction term $\mathrm{I}_{index} \cdot \Delta\mathrm{IndexTrading}_i$ is significantly positive for E-mini futures and ETFs showing that these two types of index products have a higher impact on index stock correlations than on the correlations of non-index stocks. In other words, index stock return correlations are more sensitive to changes in index trading activity. When using trading activity in all futures, $\mathrm{I}_{index} \cdot \Delta\mathrm{IndexTrading}_i$ is significant at the 10% level, confirming that futures (both floor-based and Emini) have a higher impact on index stocks than on non-index stocks.

Summarizing, the results in this section show that index trading activity also affects the correlation of non-index stocks. The impact, however, is found to be lower than for index stocks.[9]

[9]A similar comparison can be performed to test whether index trading activity has a differential impact on high beta and low beta stocks. The results, however, suggest that there is no significant difference between these two groups of stocks.

3.4.7 Are Correlations Excessive?

The previous section finds that index trading activity is positively associated with correlations. A natural question is whether the increased return comovement reflects the faster incorporation of common information or "excessive" price movements due to non-fundamental shocks. The answer to this question depends on whether demand shocks consist of pure noise or contain some information about the economy that the market has not yet incorporated into stock prices. In this section, to distinguish between these two possibilities, we follow the idea of Da and Shive (2012) and examine the relationship between return reversals and correlations.

In the first step, for every month t, we run the following regression for all stocks in the S&P 500 index

$$r^j_{it} = a^j_t + b^j_t r^j_{i-1,t} + \varepsilon^j_{it},$$

where r^j_{it} denotes stock j's return on day i in a given month t. The coefficient b^j_t measures stock j's return today relative to its return on the previous day. In the presence of a non-fundamental shock, the stock price might overshoot and then revert back to a level that is in line with its fundamentals on the next day. Therefore, a high level of return reversals indicates that returns are often driven by non-fundamental shocks. We then sum the b^j_t of all stocks weighted according to their market capitalization to obtain an aggregate measure of return reversals for a given month t:

$$b_t = \Sigma^{500}_{j=1} w_{jt} b^j_t \ .$$

Equipped with this measure of return reversals, we test whether higher return reversals are associated with greater return correlations by regressing b_t on average correlations:

$$b_t = \alpha_0 + \alpha_1 \bar{\rho}_t + \alpha_2 b_{t-1} + \varepsilon_t. \tag{3.37}$$

The idea is that if α_1 is significantly negative, then a higher correlation is associated with more return reversals. In this case, correlation is likely to be excessive. As can be seen from Panel A in Table 3.7, where the regression results are reported, α_1 is significantly negative for all subperiods, except for 1982-1997 when only legacy futures were traded. The results provide support that correlations could be excessive.

[Table 3.7 about here]

In the second step, we test the conjecture that more return reversals are also related to index trading activity. We run the following regression:

$$b_t = \beta_0 + \beta_1 \text{IndexTrading}_t + \beta_2 b_{t-1} + \varepsilon_t.$$

As reported in Panel B of Table 3.7, return reversals are significantly associated with futures (sum of E-mini and floor-based futures) and ETF trading activity. For E-mini futures, the results are weaker, but the coefficient is significantly different from zero at the 10 percent level. The regressions essentially confirm that higher index trading activity is associated with more return reversals.

Summarizing, correlations are found to have a significant relationship with return reversals, suggesting that they are excessive. Examining the relation between index trading activity and return reversals, we find that index trading activity is a significant driver of return reversals.

3.5 Conclusion

In the traditional theory, stock returns are correlated due to the correlation of their fundamentals. In this paper, we challenge this view by providing theoretical and empirical evidence that index trading activity affects average stock return correlations. Our contribution is two-fold. First, we develop an equilibrium model to investigate how ETF and futures demand shocks affect equilibrium stock prices and their correlations. Second, we empirically test whether demand shocks to index products increase stock return correlations. Our results confirm that index trading activity, a proxy for demand shocks, can explain a large part of the time variation in stock return correlations.

The most striking finding in this paper is that the correlations of non-index stocks are also significantly related to index trading activity: trading in index products not only increases the correlations of index stocks but also those of non-index stocks. When comparing the impact of index product trading on index and non-index stocks, in line with our intuition, we find that the correlations of index stocks are more sensitive to changes in index trading activity than those of non-index stocks. Furthermore, we find that despite their lower dollar trading volume, ETFs have a higher impact on correlations than futures, both theoretically and empirically.

A natural extension would be to generalize our two-period model to multiple periods. A multi-period set-up has the advantage of allowing changes in stock prices (returns) to be linked with demand shocks. However, in a multi-period model with demand shocks, the coefficient of equilibrium stock prices solves a quadratic equation and has multiple equilibria. More specifically, the number of solutions of such a model would be 2^N, where N is the number of assets.

APPENDIX

3.A Derivation of Equilibrium Prices

From Section 3.3.3, to derive the equilibrium price, we insert the expressions for the optimal portfolios into the market clearing conditions. Stock investors' optimal portfolio is given by

$$\boldsymbol{X}_s^* = (\boldsymbol{C} + \gamma_s \hat{\boldsymbol{\Sigma}})^{-1}(\hat{\boldsymbol{\mu}} - \boldsymbol{P} + \frac{1}{N_s}\boldsymbol{C}\boldsymbol{Q}). \tag{3.38}$$

For futures arbitrageurs, the optimal positions in stocks and futures are respectively

$$\boldsymbol{X}_1^* = \frac{c_f \hat{\boldsymbol{\mu}}^T \boldsymbol{b} + \gamma_1 \boldsymbol{b}^T \hat{\boldsymbol{\Sigma}} \boldsymbol{b} P_f - (\gamma_1 \boldsymbol{b}^T \hat{\boldsymbol{\Sigma}} \boldsymbol{b} + c_f)\boldsymbol{P}^T \boldsymbol{b}}{c_f \boldsymbol{b}^T \boldsymbol{C} \boldsymbol{b} + (\boldsymbol{b}^T \boldsymbol{C} \boldsymbol{b} + c_f)\gamma_1 \boldsymbol{b}^T \hat{\boldsymbol{\Sigma}} \boldsymbol{b}} \boldsymbol{b}, \tag{3.39}$$

$$y^* = \frac{-(\gamma_1 \boldsymbol{b}^T \hat{\boldsymbol{\Sigma}} \boldsymbol{b} + \boldsymbol{b}^T \boldsymbol{C} \boldsymbol{b}) P_f + \boldsymbol{b}^T \boldsymbol{C} \boldsymbol{b} \hat{\boldsymbol{\mu}}^T \boldsymbol{b} + \gamma_1 \boldsymbol{b}^T \hat{\boldsymbol{\Sigma}} \boldsymbol{b} \boldsymbol{P}^T \boldsymbol{b}}{c_f \boldsymbol{b}^T \boldsymbol{C} \boldsymbol{b} + (\boldsymbol{b}^T \boldsymbol{C} \boldsymbol{b} + c_f)\gamma_1 \boldsymbol{b}^T \hat{\boldsymbol{\Sigma}} \boldsymbol{b}}. \tag{3.40}$$

For ETF providers, the optimal positions in stocks and the ETF are

$$\boldsymbol{X}_2^* = -\frac{\boldsymbol{P}^T \boldsymbol{b} - P_{etf}}{\boldsymbol{b}^T \boldsymbol{C} \boldsymbol{b} + c_{etf}} \boldsymbol{b}, \tag{3.41}$$

$$z^* = \frac{\boldsymbol{P}^T \boldsymbol{b} - P_{etf}}{\boldsymbol{b}^T \boldsymbol{C} \boldsymbol{b} + c_{etf}}. \tag{3.42}$$

Equilibrium in ETF and futures markets requires

$$N_1 y^* + q_f = 0, \tag{3.43}$$

$$N_2 z^* + q_{etf} = 0. \tag{3.44}$$

Inserting the optimal portfolio positions (3.38)-(3.42) in the market clearing conditions yields

$$\frac{-(\gamma_1 \boldsymbol{b}^T \hat{\boldsymbol{\Sigma}} \boldsymbol{b} + \boldsymbol{b}^T \boldsymbol{C} \boldsymbol{b}) P_f + \boldsymbol{b}^T \boldsymbol{C} \boldsymbol{b} \hat{\boldsymbol{\mu}}^T \boldsymbol{b} + \gamma_1 \boldsymbol{b}^T \hat{\boldsymbol{\Sigma}} \boldsymbol{b} \boldsymbol{P}^T \boldsymbol{b}}{c_f \boldsymbol{b}^T \boldsymbol{C} \boldsymbol{b} + (\boldsymbol{b}^T \boldsymbol{C} \boldsymbol{b} + c_f)\gamma_1 \boldsymbol{b}^T \hat{\boldsymbol{\Sigma}} \boldsymbol{b}} = -\frac{q_f}{N_1}, \tag{3.45}$$

$$\frac{\boldsymbol{P}^T \boldsymbol{b} - P_{etf}}{\boldsymbol{b}^T \boldsymbol{C} \boldsymbol{b} + c_{etf}} = -\frac{q_{etf}}{N_2}. \tag{3.46}$$

Using these conditions, we can express the equilibrium futures and ETF prices in terms of demand shocks and the stocks prices as

$$P_f = \frac{\frac{q_f}{N_1} a + \boldsymbol{b}^T \boldsymbol{C} \boldsymbol{b} \hat{\boldsymbol{\mu}}^T \boldsymbol{b} + \gamma_1 \boldsymbol{b}^T \hat{\boldsymbol{\Sigma}} \boldsymbol{b} \boldsymbol{P}^T \boldsymbol{b}}{\gamma_1 \boldsymbol{b}^T \hat{\boldsymbol{\Sigma}} \boldsymbol{b} + \boldsymbol{b}^T \boldsymbol{C} \boldsymbol{b}}, \tag{3.47}$$

$$P_{etf} = \boldsymbol{P}^T \boldsymbol{b} + \frac{q_{etf}}{N_2} (\boldsymbol{b}^T \boldsymbol{C} \boldsymbol{b} + c_{etf}) \,, \tag{3.48}$$

where $a \equiv c_f \boldsymbol{b}^T \boldsymbol{C} \boldsymbol{b} + (\boldsymbol{b}^T \boldsymbol{C} \boldsymbol{b} + c_f) \gamma_1 \boldsymbol{b}^T \hat{\boldsymbol{\Sigma}} \boldsymbol{b}$. The optimal investment in stocks can then be expressed solely in terms of demand shocks and stock prices,

$$\begin{aligned} \boldsymbol{X}_1^* &= \frac{c_f \hat{\boldsymbol{\mu}}^T \boldsymbol{b} + \gamma_1 \boldsymbol{b}^T \hat{\boldsymbol{\Sigma}} \boldsymbol{b} \frac{\frac{q_f}{N_1} a + \boldsymbol{b}^T \boldsymbol{C} \boldsymbol{b} \hat{\boldsymbol{\mu}}^T \boldsymbol{b} + \gamma_1 \boldsymbol{b}^T \hat{\boldsymbol{\Sigma}} \boldsymbol{b} \boldsymbol{P}^T \boldsymbol{b}}{\gamma_1 \boldsymbol{b}^T \hat{\boldsymbol{\Sigma}} \boldsymbol{b} + \boldsymbol{b}^T \boldsymbol{C} \boldsymbol{b}} - (\gamma_1 \boldsymbol{b}^T \hat{\boldsymbol{\Sigma}} \boldsymbol{b} + c_f) \boldsymbol{P}^T \boldsymbol{b}}{a} \boldsymbol{b} \\ &= \frac{\gamma_1 \boldsymbol{b}^T \hat{\boldsymbol{\Sigma}} \boldsymbol{b}}{\gamma_1 \boldsymbol{b}^T \hat{\boldsymbol{\Sigma}} \boldsymbol{b} + \boldsymbol{b}^T \boldsymbol{C} \boldsymbol{b}} \frac{q_f}{N_1} \boldsymbol{b} + \frac{\left(c_f + \frac{\gamma_1 \boldsymbol{b}^T \hat{\boldsymbol{\Sigma}} \boldsymbol{b} \boldsymbol{b}^T \boldsymbol{C} \boldsymbol{b}}{\gamma_1 \boldsymbol{b}^T \hat{\boldsymbol{\Sigma}} \boldsymbol{b} + \boldsymbol{b}^T \boldsymbol{C} \boldsymbol{b}} \right) \boldsymbol{b} \boldsymbol{b}^T}{a} (\hat{\boldsymbol{\mu}} - \boldsymbol{P}) \,, \end{aligned} \tag{3.49}$$

$$\boldsymbol{X}_2^* = \frac{q_{etf}}{N_2} \boldsymbol{b} \,. \tag{3.50}$$

Inserting the optimal stock investment into the stock market equilibrium condition, we have

$$\begin{aligned} & N_s (\boldsymbol{C} + \gamma_s \hat{\boldsymbol{\Sigma}})^{-1} (\hat{\boldsymbol{\mu}} - \boldsymbol{P} + \frac{1}{N_s} \boldsymbol{C} \boldsymbol{Q}) + N_1 \frac{\gamma_1 \boldsymbol{b}^T \hat{\boldsymbol{\Sigma}} \boldsymbol{b}}{\gamma_1 \boldsymbol{b}^T \hat{\boldsymbol{\Sigma}} \boldsymbol{b} + \boldsymbol{b}^T \boldsymbol{C} \boldsymbol{b}} \frac{q_f}{N_1} \boldsymbol{b} \\ & + \frac{N_1 \left(c_f + \frac{\gamma_1 \boldsymbol{b}^T \hat{\boldsymbol{\Sigma}} \boldsymbol{b} \boldsymbol{b}^T \boldsymbol{C} \boldsymbol{b}}{\gamma_1 \boldsymbol{b}^T \hat{\boldsymbol{\Sigma}} \boldsymbol{b} + \boldsymbol{b}^T \boldsymbol{C} \boldsymbol{b}} \right) \boldsymbol{b} \boldsymbol{b}^T}{a} (\hat{\boldsymbol{\mu}} - \boldsymbol{P}) + N_2 \frac{q_{etf}}{N_2} \boldsymbol{b} = \boldsymbol{Q} \,. \end{aligned} \tag{3.51}$$

Simplifying and defining $\alpha = \frac{\gamma_1 \boldsymbol{b}^T \hat{\boldsymbol{\Sigma}} \boldsymbol{b}}{\gamma_1 \boldsymbol{b}^T \hat{\boldsymbol{\Sigma}} \boldsymbol{b} + \boldsymbol{b}^T \boldsymbol{C} \boldsymbol{b}}$ and $\lambda = \frac{N_1}{a} \left(c_f + \frac{\gamma_1 \boldsymbol{b}^T \hat{\boldsymbol{\Sigma}} \boldsymbol{b} \boldsymbol{b}^T \boldsymbol{C} \boldsymbol{b}}{\gamma_1 \boldsymbol{b}^T \hat{\boldsymbol{\Sigma}} \boldsymbol{b} + \boldsymbol{b}^T \boldsymbol{C} \boldsymbol{b}} \right)$, we have

$$N_s (\boldsymbol{C} + \gamma_s \hat{\boldsymbol{\Sigma}})^{-1} (\hat{\boldsymbol{\mu}} - \boldsymbol{P} + \frac{1}{N_s} \boldsymbol{C} \boldsymbol{Q}) + N_1 \alpha \frac{q_f}{N_1} \boldsymbol{b} + \lambda \boldsymbol{b} \boldsymbol{b}^T (\hat{\boldsymbol{\mu}} - \boldsymbol{P}) + N_2 \frac{q_{etf}}{N_2} \boldsymbol{b} = \boldsymbol{Q} \,. \tag{3.52}$$

Solving, equilibrium stock prices are given by

$$\boldsymbol{P} = \hat{\boldsymbol{\mu}} + \alpha \boldsymbol{D}^{-1} \boldsymbol{b} q_f + \boldsymbol{D}^{-1} \boldsymbol{b} q_{etf} + \boldsymbol{D}^{-1} \left((\boldsymbol{C} + \gamma_s \hat{\boldsymbol{\Sigma}})^{-1} \boldsymbol{C} - \boldsymbol{I} \right) \boldsymbol{Q}, \tag{3.53}$$

where $\boldsymbol{D} = N_s (\boldsymbol{C} + \gamma_s \hat{\boldsymbol{\Sigma}})^{-1} + \lambda \boldsymbol{b} \boldsymbol{b}^T$. The equilibrium futures and ETF prices can be obtained by inserting equilibrium stock prices into equations (3.47) and (3.48)

$$
\begin{aligned}
P_f &= \frac{\frac{q_f}{N_1}a + \boldsymbol{b}^T\boldsymbol{C}\boldsymbol{b}\hat{\boldsymbol{\mu}}^T\boldsymbol{b}}{\gamma_1\boldsymbol{b}^T\hat{\boldsymbol{\Sigma}}\boldsymbol{b} + \boldsymbol{b}^T\boldsymbol{C}\boldsymbol{b}} + \gamma_1\alpha\boldsymbol{b}^T\left(\hat{\boldsymbol{\mu}} + \alpha\boldsymbol{D}^{-1}\boldsymbol{b}q_f + \boldsymbol{D}^{-1}\boldsymbol{b}q_{etf} + \boldsymbol{D}^{-1}\left((\boldsymbol{C}+\gamma_s\hat{\boldsymbol{\Sigma}})^{-1}\boldsymbol{C} - \boldsymbol{I}\right)\boldsymbol{Q}\right) \\
P_{etf} &= \boldsymbol{b}^T\left(\hat{\boldsymbol{\mu}} + \alpha\boldsymbol{D}^{-1}\boldsymbol{b}q_f + \boldsymbol{D}^{-1}\boldsymbol{b}q_{etf} + \boldsymbol{D}^{-1}\left((\boldsymbol{C}+\gamma_s\hat{\boldsymbol{\Sigma}})^{-1}\boldsymbol{C} - \boldsymbol{I}\right)\boldsymbol{Q}\right) + \frac{q_{etf}}{N_2}(\boldsymbol{b}^T\boldsymbol{C}\boldsymbol{b} + c_{etf}) .
\end{aligned}
$$

3.B Impact of Demand Shocks on Prices in a Special Case

In this section, we prove that in the special case when stocks have independent liquidity ($\boldsymbol{C}$ is a diagonal matrix) and investors have independent posterior beliefs ($\hat{\boldsymbol{\Sigma}}$ is a diagonal matrix), demand shocks only impact index stocks. In other words, we prove that $\boldsymbol{D}^{-1}\boldsymbol{b}$ has elements larger than or equal to zero for stocks included in the index and zero entries for non-index stocks when $\boldsymbol{C}$ and $\hat{\boldsymbol{\Sigma}}$ are diagonal matrices.

Let $\boldsymbol{B}$ denote the matrix $\lambda\boldsymbol{b}\boldsymbol{b}^T$. Since $\boldsymbol{b}$ only has non-zero entries for the first k stocks which are included in the index, $\boldsymbol{B}$ can be written as a block matrix: $\boldsymbol{B} = \begin{pmatrix} \boldsymbol{B_1} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{0} \end{pmatrix}$, where $\boldsymbol{B_1}$ is a $k \times k$ matrix. It can be proved that $\boldsymbol{B}$ is a positive semi-definite matrix.

Define $\boldsymbol{A} = N_s(\boldsymbol{C} + \gamma_s\hat{\boldsymbol{\Sigma}})^{-1}$. Since $\boldsymbol{C}$ and $\hat{\boldsymbol{\Sigma}}$ are positive definite diagonal matrices, $\boldsymbol{A}$ is also a positive definite diagonal matrix. We can decompose $\boldsymbol{A}$ into two diagonal matrices as $\begin{pmatrix} \boldsymbol{A_1} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{A_2} \end{pmatrix}$, where $\boldsymbol{A_1}$ and $\boldsymbol{A_2}$ are of dimension $k \times k$ and $N-k \times N-k$, respectively. $\boldsymbol{D}^{-1}\boldsymbol{b}$ can be written in terms of $\boldsymbol{A}$ and $\boldsymbol{B}$:

$$
\boldsymbol{D}^{-1}\boldsymbol{b} = (\boldsymbol{A}+\boldsymbol{B})^{-1}\boldsymbol{b} = \begin{pmatrix} (\boldsymbol{B_1}+\boldsymbol{A_1})^{-1} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{A_2}^{-1} \end{pmatrix}\boldsymbol{b} . \tag{3.54}
$$

Because the $(k+1)$th to Nth elements of $\boldsymbol{b}$ are zero (these are the weights for non-index stocks), from equation (3.54), it follows that $\boldsymbol{D}^{-1}\boldsymbol{b}$ also has zero entries for elements $k+1$ to N which are the coefficients for non-index stocks.

We need to prove that the first k entries of $\boldsymbol{D}^{-1}\boldsymbol{b}$ are larger than or equal to zero. Let $\bar{\boldsymbol{D}}$ denote the adjugate matrix of $\boldsymbol{D}$, the inverse of $\boldsymbol{D}$ can be written in terms of the adjugate matrix as follows

$$
\boldsymbol{D}^{-1} = \frac{1}{\det(\boldsymbol{D})}\bar{\boldsymbol{D}} , \tag{3.55}
$$

where $\bar{D}_{ij} = (-1)^{i+j}\det\left((\boldsymbol{D}_{k,h})_{1\le k\ne j\le N, 1\le h\ne i\le N}\right)$.

The ith element of $\boldsymbol{D}^{-1}\boldsymbol{b}$ can also be computed using the adjugate matrix as follows

$$
\begin{aligned}
\frac{1}{\det(\boldsymbol{D})}\sum_{j=1}^{N}\bar{\boldsymbol{D}}_{ij}b_j &= \frac{1}{\det(\boldsymbol{D})}\det\begin{pmatrix} B_{11}+A_{11} & B_{12} & .. & B_{1,i-1} & b_1 & B_{1,i+1}... & B_{1n} \\ B_{21} & B_{22}+A_{22} & .. & B_{2,i-1} & b_2 & B_{2,i+1}... & B_{2n} \\ ... & & & & & & \\ B_{n1} & B_{n2} & .. & B_{n,i-1} & b_n & B_{n,i+1}... & B_{nn}+A_{nn} \end{pmatrix} \\
&= \frac{1}{\det(\boldsymbol{D})\lambda b_i}\det\begin{pmatrix} B_{11}+A_{11} & B_{12} & .. & B_{1,i-1} & B_{1,i} & B_{1,i+1}... & B_{1n} \\ B_{21} & B_{22}+A_{22} & .. & B_{2,i-1} & B_{2,i} & B_{2,i+1}... & B_{2n} \\ ... & & & & & & \\ B_{n1} & B_{n2} & .. & B_{n,i-1} & B_{n,i} & B_{n,i+1}... & B_{nn}+A_{nn} \end{pmatrix} \\
&= \frac{1}{\det(\boldsymbol{D})\lambda b_i}\det\left(\boldsymbol{B}^T+\begin{pmatrix} A_{11} & & & & & & \\ ... & A_{22} & ... & & & & \\ ... & & & & & & \\ ... & & A_{i-1} & ... & & & \\ ... & & & 0 & ... & & \\ ... & & & & A_{i+1} & .. & \\ .. & & & & & & \\ ... & & & & & & A_{nn} \end{pmatrix}\right) \geq 0\,.
\end{aligned}
$$

The last inequality follows from the fact that the determinant of a positive semi-definite matrix is larger than or equal to zero. Hence, $\boldsymbol{D}^{-1}\boldsymbol{b}$ has positive or zero entries for index stocks.

Panel A

	Sample period	Obs.	Mean	SD	Min	Max	Skewness	Excess kurtosis	ADF test
ρ_t (S&P)	04.1982-12.2012	369	0.299	0.135	0.064	0.824	0.860	0.860	-5.273
$\bar{\rho}_t$ (S&P)	04.1982-12.2012	369	0.316	0.164	0.064	1.171	1.432	3.328	-5.453
ρ_t (high beta)	04.1982-12.2012	369	0.252	0.144	0.031	0.775	1.170	1.049	-5.557
$\bar{\rho}_t$ (high beta)	04.1982-12.2012	369	0.266	0.171	0.031	1.033	1.565	2.706	-5.641
ρ_t (low beta)	04.1982-12.2012	369	0.154	0.127	0.020	0.763	1.675	2.948	-5.535
$\bar{\rho}_t$ (low beta)	04.1982-12.2012	369	0.164	0.141	0.024	1.000	2.066	5.518	-5.671
SPVol_t	04.1982-12.2012	369	0.044	0.025	0.009	0.273	3.732	23.982	-7.910
CurrentReturn_t(%)	04.1982-12.2012	369	0.598	4.417	-23.077	14.016	-0.629	2.748	-14.412
$\Delta\text{3MonthTbill}_t$(%)	04.1982-12.2012	369	-0.034	0.274	-2.669	0.587	-3.101	23.480	-7.557
CreditSpread_t(%)	04.1982-12.2012	369	1.075	0.462	0.550	3.380	2.105	5.770	-4.030
$\Delta\text{IndustrialProduction}_t$	04.1982-12.2012	369	0.009	0.018	-0.069	0.047	-1.520	3.826	-3.990
Inflation_t	04.1982-12.2012	369	0.001	0.001	-0.008	0.005	-1.073	6.693	-7.740
$\text{PolicyUncertainty}_t$	01.1985-12.2012	336	106.710	32.176	57.206	245.126	1.029	0.963	-4.476
Sentiment_t	04.1982-12.2010	345	-0.028	0.342	-1.270	1.333	-0.161	1.958	-4.075

Panel B

	Sample period	Obs.	Mean	SD	Min	Max	Skewness	Excess kurtosis	ADF test
FuturesRatio_t	04.1982-12.2012	369	1.168	0.315	0.075	2.130	0.207	0.247	-2.827
LegacyRatio_t	04.1982-08.1997	185	1.244	0.333	0.075	2.130	0.151	0.193	-3.361
EminiRatio_t	09.1997-12.2012	184	0.696	0.445	0.032	1.631	-0.197	-1.164	-1.998
ETFRatio_t	02.1993-12.2012	240	0.100	0.099	0.000	0.377	0.725	-0.694	-3.310
$\Delta\text{FuturesRatio}_t$	04.1982-12.2012	369	0.006	0.140	-0.462	0.631	0.362	1.371	-13.402
$\Delta\text{LegacyRatio}_t$	04.1982-12.2012	185	0.010	0.158	-0.462	0.631	0.496	1.364	-9.637
$\Delta\text{EminiRatio}_t$	09.1997-12.2012	184	0.009	0.096	-0.280	0.378	0.171	1.583	-10.297
$\Delta\text{ETFRatio}_t$	02.1993-12.2012	240	0.001	0.022	-0.090	0.096	0.772	4.226	-7.221

TABLE 3.1: *Summary Statistics*

The table presents summary statistics of the variables used in the regressions. The summary statistics include the sample period, number of observations, mean, standard deviation (SD), minimum, maximum, skewness, excess kurtosis and the value of the augmented Dicky-Fuller (ADF) test. The lag of the ADF test is chosen according to the Schwarz information criterion. Panel A shows summary statistics for independent and dependent variables excluding trading ratios. ρ_t is the average Pearson correlation of stocks. $\bar{\rho}_t$ denotes the Fisher transformed correlations. SPVol_t is the monthly realized volatility of the S&P 500 index. CurrentReturn_t(%) is the current month S&P 500 return measured in percentage points. $\Delta\text{3MonthTbill}_t$(%) measures the difference between the 3-month treasury bill rate and its value for the previous month. CreditSpread_t(%) is the yield difference between AAA bonds and BAA bonds. $\Delta\text{IndustrialProduction}_t$ is the annual industrial production growth compared with the same month in the previous year. Inflation_t is computed as the logarithm of the ratio of the values of the CPI at t and $t-1$. $\text{PolicyUncertainty}_t$ is the value of policy uncertainty index of Baker, Bloom and Davis. Panel B reports the summary statistics for the trading ratios. Index trading ratios are defined as the index derivatives dollar trading volume scaled by the aggregate dollar trading volume of S&P 500 stocks. For example, FuturesRatio_t is the futures dollar trading volume including floor-based as well as Emini futures divided by the aggregate stock trading volume. Δ indicates that trading ratios are fitted to ARIMA models. For the variables with Δ, the summary statistics are provided for the residual from the ARIMA model .

	$\bar{\rho}_t$	$SPVol_t$	ΔFutures	ΔETF_t	$\Delta Emini_t$	$CurrentReturn_t$	$\Delta 3MonthTbill_t$	$CreditSpread_t$	$\Delta IndustrialProduction_t$	$Inflation_t$	$PolicyUncertainty_t$	$Sentiment_t$
$\bar{\rho}_t$	1											
$SPVol_t$	0.616	1										
$\Delta FuturesRatio_t$	0.173	-0.281	1									
$\Delta ETFRatio_t$	0.420	0.232	0.488	1								
$\Delta EminiRatio_t$	0.285	0.077	0.707	0.686	1							
$CurrentReturn_t$(%)	-0.263	-0.281	-0.209	-0.377	-0.285	1						
$\Delta 3MonthTbill_t$(%)	-0.132	-0.245	-0.003	-0.169	-0.081	-0.020	1					
$CreditSpread_t$(%)	0.362	0.382	0.041	-0.126	-0.108	0.033	-0.219	1				
$\Delta IndustrialProduction_t$	-0.217	-0.169	-0.070	0.045	-0.023	0.008	0.086	-0.450	1			
$Inflation_t$	0.362	-0.229	0.116	-0.126	0.113	-0.056	0.083	-0.130	0.101	1		
$PolicyUncertainty_t$	0.364	0.249	-0.126	-0.020	-0.012	-0.091	-0.061	-0.020	-0.142	-0.123	1	
$Sentiment_t$	-0.210	-0.140	-0.015	0.047	-0.005	-0.066	0.019	-0.287	0.213	0.138	-0.155	1

TABLE 3.2: *Correlations*

This table reports the correlations between variables that are used in the regressions. For some variables that are not available for the whole sample period from 1982 to 2012, the correlation is calculated for the subsample when they are available. For example, to compute the correlation of $\Delta ETFRatio_t$ with other variables, a subsample of 1993-2012 is used. Similarly, the correlation of $\Delta EminiRatio_t$ with other variables is computed for the period 1997-2012.

	All Futures 1982-2012		*Legacy 1982-1997*		*Emini 1997-2012*		*ETFs 1993-2012*		
ΔAllFuturesRatio$_t$		**0.241**							**0.176**
		(3.805)							**(4.650)**
ΔLegacyFuturesRatio$_t$				**0.103**					
				(2.123)					
ΔEminiFuturesRatio$_t$						**0.612**			
						(7.407)			
ΔETFRatio$_t$								**3.748**	**3.295**
								(7.447)	**(6.544)**
SPVol$_t$	**2.648**	**2.824**	**3.703**	**3.889**	**1.559**	**1.307**	**1.839**	**0.819**	**0.882**
	(6.185)	**(6.352)**	**(11.367)**	**(14.190)**	**(2.405)**	**(2.440)**	**(2.971)**	**(2.272)**	**(2.637)**
CurrentReturn$_t$	**-0.527**	**-0.370**	-0.100	-0.022	**-0.752**	**-0.444**	**-0.741**	-0.142	-0.061
	(-3.293)	**(-2.348)**	(-0.443)	(-0.093)	**(-3.717)**	**(-2.173)**	**(-4.173)**	(-0.783)	(-0.345)
Δ3MonthTbill$_t$	-0.002	-0.000	0.014	0.011	-0.007	0.008	-0.010	0.039	0.038
	(-0.089)	(-0.002)	(0.663)	(0.592)	(-0.079)	(0.098)	(-0.131)	(0.772)	(0.729)
CreditSpread$_t$	0.019	0.010	0.006	-0.002	0.032	0.051	0.033	**0.074**	**0.079**
	(0.926)	(0.453)	(0.388)	(-0.177)	(0.855)	(1.590)	(0.867)	**(2.755)**	**(3.053)**
ΔIndustrialProduction$_t$	0.122	0.160	-0.449	-0.386	0.678	0.840	0.624	0.469	0.572
	(0.317)	(0.373)	(-1.080)	(-0.853)	(1.150)	(1.684)	(1.111)	(0.952)	(1.279)
Inflation$_t$	**7.904**	6.037	**16.733**	**13.896**	2.254	-1.046	4.125	2.019	1.847
	(2.404)	(1.874)	**(3.172)**	**(2.660)**	(0.476)	(-0.243)	(0.979)	(0.654)	(0.636)
PolicyUncertainty$_t$	-	-	-	-	0.092	0.085	0.080	**0.082**	**0.079**
	-	-	-	-	(1.652)	(1.685)	(1.486)	**(2.483)**	**(2.356)**
$\bar{\rho}_{t-1}$	**0.419**	**0.434**	**0.286**	**0.294**	**0.383**	**0.423**	**0.420**	**0.478**	**0.474**
	(7.750)	**(8.333)**	**(5.784)**	**(5.652)**	**(3.881)**	**(5.066)**	**(5.108)**	**(9.889)**	**(10.226)**
Adjusted R^2	0.539	0.578	0.671	0.684	0.481	0.574	0.557	0.724	0.734

TABLE 3.3: *Regression results for S&P 500 stocks.*

The table reports the estimated coefficients from regressing the average correlation of index stocks on different explanatory variables including index trading activity, monthly realized volatility, the current month market return, the 3-month treasury bill rate, the credit spread, the growth in industrial production, inflation and policy uncertainty. Policy uncertainty is only available since 1985, thus it is only employed for the periods 1993-2012 and 1997-2012. The t-statistics shown in parentheses are based on Newey-West standard errors with 12 lags. The coefficients for sentiment are not reported, as this variable is only available until 2010. We run the regression for the subperiod until 2010 and find that sentiment is not significant. For brevity, the results including sentiment are not reported.

	All Futures 1982-2012		*Legacy 1982-1997*		*Emini 1997-2012*		*ETFs 1993-2012*		
ΔAllFuturesRatio$_t$		**0.173**							**0.183**
		(2.428)							**(2.743)**
ΔLegacyFuturesRatio$_t$				0.013					
				(0.281)					
ΔEminiFuturesRatio$_t$						**0.615**			
						(7.921)			
ΔETFRatio$_t$								**3.053**	**2.583**
								(6.312)	**(5.136)**
ΔSPVol$_t$	**2.321**	**2.449**	**2.849**	**2.874**	**1.188**	**0.978**	**1.373**	0.601	**0.661**
	(8.142)	**(7.603)**	**(10.560)**	**(10.434)**	**(2.415)**	**(2.439)**	**(3.044)**	(1.742)	**(2.001)**
CurrentReturn$_t$	**-0.482**	**-0.344**	**-0.543**	**-0.532**	-0.278	0.029	-0.367	0.118	0.202
	(-3.075)	**(-2.117)**	**(-3.537)**	**(-3.103)**	(-1.217)	(0.122)	(-1.778)	(0.560)	(0.947)
Δ3MonthTbill$_t$	0.027	0.028	0.027	0.027	0.017	0.029	0.012	0.049	0.048
	(1.103)	(1.171)	(1.649)	(1.636)	(0.253)	(0.485)	(0.208)	(1.226)	(1.161)
CreditSpread$_t$	-0.005	-0.012	-0.036	**-0.037**	0.068	**0.087**	**0.072**	**0.104**	**0.109**
	(-0.270)	(-0.525)	(-1.835)	**(-2.013)**	(1.951)	**(3.050)**	**(2.002)**	**(3.645)**	**(3.962)**
ΔIndustrialProduction$_t$	-0.338	-0.297	0.261	0.268	0.514	0.730	0.433	0.364	0.466
	(-0.654)	(-0.582)	(0.537)	(0.542)	(0.694)	(1.223)	(0.613)	(0.612)	(0.833)
Inflation$_t$	2.486	1.263	6.036	5.663	0.884	-2.465	2.107	0.394	0.216
	(0.838)	(0.410)	(1.052)	(0.954)	(0.250)	(-0.773)	(0.592)	(0.116)	(0.067)
PolicyUncertainty$_t$	-	-	-	-	**0.122**	**0.113**	**0.111**	**0.108**	**0.106**
	-	-	-	-	**(2.632)**	**(2.615)**	**(2.382)**	**(3.759)**	**(3.537)**
$\bar{\rho}_{t-1}$	**0.576**	**0.590**	**0.181**	**0.183**	**0.406**	**0.445**	**0.437**	**0.487**	**0.483**
	(9.362)	**(9.477)**	**(2.916)**	**(3.129)**	**(7.325)**	**(7.932)**	**(9.378)**	**(12.120)**	**(11.740)**
Adjusted R^2	0.583	0.601	0.554	0.551	0.568	0.653	0.631	0.729	0.739

TABLE 3.4: *Regression results for high beta non-S&P 500 stocks.*

The table reports the estimated coefficients from regressing the average correlation of high beta stocks on different explanatory variables. The explanatory variables are the same as in the regression for index stocks. We include index trading activity, monthly realized volatility, the current month market return, the 3-month treasury bill rate, the credit spread, the growth in industrial production, inflation and policy uncertainty. Policy uncertainty is only available since 1985, thus it is only employed for the periods 1993-2012 and 1997-2012. The t-statistics shown in parentheses are based on Newey-West standard errors with 12 lags. The coefficients for sentiment are not reported, as this variable is only available until 2010. We run the regression for the subperiod until 2010 and find that sentiment is not significant. For brevity, the results including sentiment are not reported.

	All Futures 1982-2012		*Legacy 1982-1997*		*Emini 1997-2012*		*ETFs 1993-2012*		
ΔAllFuturesRatio$_t$		**0.113**							0.071
		(2.424)							(1.688)
ΔLegacyFuturesRatio$_t$				-0.001					
				(0.098)					
ΔEminiFuturesRatio$_t$						**0.428**			
						(6.486)			
ΔETFRatio$_t$								**2.447**	**2.264**
								(5.349)	**(4.742)**
ΔSPVol$_t$	**1.466**	**1.550**	**2.112**	**2.108**	0.674	0.518	**0.780**	0.156	0.179
	(5.749)	**(5.284)**	**(8.381)**	**(7.980)**	(1.760)	(1.662)	**(2.285)**	(0.749)	(0.882)
CurrentReturn$_t$	**-0.302**	-0.210	**-0.273**	**-0.274**	-0.225	-0.000	-0.263	0.140	0.173
	(-2.420)	(-1.692)	**(-3.113)**	**(-3.085)**	(-1.088)	(-0.001)	(-1.456)	(0.724)	(0.886)
Δ3MonthTbill$_t$	0.014	0.015	0.014	0.014	0.013	0.022	0.003	0.034	0.034
	(0.770)	(0.840)	(1.194)	(1.195)	(0.234)	(0.434)	(0.072)	(1.030)	(1.005)
CreditSpread$_t$	-0.010	-0.015	**-0.033**	**-0.033**	0.038	**0.054**	0.039	**0.072**	**0.074**
	(-0.822)	(-1.007)	**(-2.980)**	**(-2.995)**	(1.363)	**(2.271)**	(1.476)	**(4.136)**	**(4.363)**
ΔIndustrialProduction$_t$	-0.267	-0.241	0.000	-0.000	0.477	0.641	0.392	0.365	0.409
	(-0.765)	(-0.669)	(0.013)	(-0.002)	(0.945)	(1.481)	(0.865)	(0.936)	(1.098)
Inflation$_t$	1.662	0.815	3.108	3.157	-0.066	-2.592	0.895	-0.685	-0.761
	(0.700)	(0.333)	(0.971)	(0.983)	(-0.235)	(-0.884)	(0.315)	(-0.244)	(-0.275)
PolicyUncertainty$_t$	-	-	-	-	0.093	0.090	0.082	**0.086**	**0.085**
	-	-	-	-	(1.904)	(1.884)	(1.768)	**(2.565)**	**(2.494)**
$\bar{\rho}_{t-1}$	**0.667**	**0.686**	**0.150**	**0.150**	**0.507**	**0.520**	**0.551**	**0.565**	**0.560**
	(12.412)	**(13.025)**	**(2.858)**	**(2.972)**	**(7.466)**	**(8.054)**	**(9.869)**	**(13.492)**	**(13.546)**
Adjusted R^2	0.641	0.652	0.657	0.6552	0.585	0.647	0.655	0.751	0.752

TABLE 3.5: *Regression results for low beta non-S&P 500 stocks.*

The table reports the estimated coefficients from regressing the average correlation of low beta stocks on different explanatory variables. The explanatory variables are the same as in the regression for index stocks. We include index trading activity, monthly realized volatility, the current month market return, the 3-month treasury bill rate, the credit spread, the growth in industrial production, inflation and policy uncertainty. Policy uncertainty is only available since 1985, thus it is only employed for the periods 1993-2012 and 1997-2012. The t-statistics shown in parentheses are based on Newey-West standard errors with 12 lags. The coefficients for sentiment are not reported, as this variable is only available until 2010. We run the regression for the subperiod until 2010 and find that sentiment is not significant. For brevity, the results including sentiment are not reported.

	All Futures 1982-2012	*Emini* 1997-2012	*ETFs* 1993-2012
ΔIndexTrading	**0.100**	**0.473**	**2.322**
	(2.622)	**(7.555)**	**(7.361)**
I_{index}	**0.103**	**0.066**	**0.071**
	(5.939)	**(3.638)**	**(4.534)**
$\mathrm{I}_{index} \cdot$ ΔIndexTrading	0.107	**0.252**	**1.533**
	(1.902)	**(2.194)**	**(2.952)**
SPVol	**2.913**	**0.705**	**0.743**
	(10.634)	**(2.452)**	**(2.550)**
CurrentReturn	**-0.281**	-0.036	0.191
	(-2.525)	(-0.265)	(1.710)
Δ3MonthTbill	0.033	0.061	0.060
	(1.674)	(1.479)	(1.662)
CreditSpread	**0.054**	**0.160**	**0.185**
	(2.180)	**(4.797)**	**(5.238)**
ΔIndustrialProduction	-0.200	1.440	1.232
	(-0.283)	(1.175)	(1.808)
Inflation	-4.914	-5.221	-2.228
	(-1.208)	(-1.443)	(-0.583)
PolicyUncertainty	-	**0.172**	**0.182**
	-	**(4.952)**	**(5.004)**
Adjusted R^2	0.383	0.544	0.589

TABLE 3.6: *Comparison of the impact of index trading activity on the correlations of index and non-index stocks.*

This table reports the results of pooled regressions of the correlations between index stocks, low beta stocks and high beta stocks on a number of explanatory variables. The dependent variable is the monthly average correlation of index stocks, high beta stocks and low beta stocks. In addition to the explanatory variables in Table 3.3-3.5, we include a dummy variable I_{index} indicating whether the correlation relates to index stocks or not and its interaction with index trading $\mathrm{I}_{index} \cdot$ ΔIndexTrading. For the period 1982-2012, ΔIndexTrading includes trading activity in all futures contracts. For 1997-2012, it includes Emini futures trading activity. For the subperiod 1993-2012, ΔIndexTrading includes ETF trading activity. The t-statistics shown in parentheses are based on Newey-West standard errors with 0 lag to correct for heteroscedasticity. Since we pool the correlations of index stocks, low beta stocks and high beta stocks in one regression, there is no need to adjust for autocorrelation.

Panel A

	1982-2012	*1982-1997*	*1997-2012*	*1993-2012*
$\bar{\rho}_t$	**-0.121**	-0.032	**-0.141**	**-0.138**
	(-4.737)	(-0.896)	**(-5.727)**	**(-6.131)**
b_{t-1}	0.060	0.028	0.028	-0.060
	(0.713)	(0.279)	(0.380)	(-0.537)
Adjusted R^2	0.079	0	0.106	0.117

Panel B

	All Futures *1982-2012*	*Legacy* *1982-1997*	*Emini* *1997-2012*	*ETFs* *1993-2012*
ΔIndexTrading	**-0.053**	-0.039	-0.091	**-0.645**
	(-2.573)	(-1.428)	(-1.823)	**(-2.794)**
b_{t-1}	0.113	0.0522	0.028	0.0398
	(1.308)	(0.509)	(0.380)	(0.310)
Adjusted R^2	0.020	0	0.0311	0.005

TABLE 3.7: *Testing whether correlations are excessive.*

This table shows the results of regressing return reversals on correlations (Panel A) and index trading activity (Panel B). The dependent variable in both regressions is the monthly weighted average of return reversals. The lag of the dependent variable b_{t-1} is included in the regression to correct for autocorrelation.

Chapter 4

The Impact of Liquidity on Information Acquisition

Abstract

This paper extends the information acquisition problem considered in Van Nieuwerburgh and Veldkamp (2010) to a more realistic setting with transaction costs, including both proportional and quadratic transaction costs. Our findings differ from those obtained without transaction costs. As an asset's transaction costs rise, it becomes less attractive for investors to learn about it. Investors' decision about which assets to learn depends on their initial holdings. Moreover, transaction costs might change investors' information acquisition choice from specialized learning to generalized learning. Hence, in addition to assets' Sharpe ratio, investors' initial asset holdings and assets' liquidity play an important role in determining optimal information acquisition strategies.

The author is extremely grateful to Maria Putintseva for her suggestions and help for checking the computations.

4.1 Introduction

Many players in financial markets, such as mutual fund and hedge fund managers, actively search for superior information regarding the value of traded assets in order to profit from such information. Professional investors spend substantial amounts of time and resources acquiring information about individual companies, industries, and the macroeconomy to sharpen their investment decisions. However, although investors might acquire substantial economic and financial information regarding a traded asset, investment opportunities are commonly subject to a type of uncertainty that resolves only ex post: imperfect information regarding the investment payoff. The present paper examines this type of uncertainty and assumes that investors can exert effort to collect information about assets' payoffs.

Van Nieuwerburgh and Veldkamp (2010) assume that investors can acquire information before they make their portfolio choice and develop a model in which portfolio and information decisions are jointly determined. In their model, information can be used to make more precise predictions about asset payoffs, i.e. to reduce the conditional variance of such payoffs. When the possibility of learning exists, investors with different objective functions and learning technologies make different information and portfolio decisions.

However, the Van Nieuwerburgh and Veldkamp (2010) model is built on the assumption of a perfectly liquid market. In their model, assets can be purchased and sold immediately without any cost and in any quantity. Such an assumption is not consistent with actual markets in which assets are costly to trade. In this paper, we investigate the impact of liquidity on investors' information acquisition decision. It is well-established that transaction costs are a significant factor in determining the trading behavior of financial market participants, optimal asset allocations, and assets' equilibrium prices. However, the direction and magnitude of the impact of transaction costs on information acquisition have not previously been analyzed in the literature.

Our study's contribution is to introduce transaction costs into a model based on Van Nieuwerburgh and Veldkamp (2010) and to characterize the impact of transaction costs on optimal information acquisition. Because transaction costs affect investors' optimal investment strategies, optimal information allocation is also affected. For example, investors might take a more aggressive position in an asset when that asset's transaction costs drop. Accordingly, investors might collect more information on these assets that have low transaction costs because higher benefits can now be obtained from learning about them than from learning about assets with high transaction costs.

We consider two types of transaction costs in our model: proportional and quadratic transaction costs. Proportional transaction costs have been used in numerous studies (see Constantinides, Jackwerth, and Perrakis (2009) for example). Typically, there is a no-trade region, and investors that begin in the no-trade region simply do not trade; however, any allocation outside the no-trade region causes investors to trade until they reach the nearest boundary of the no-trade region. We consider a situation with two risky assets and one bundle. Although we do not have a closed-form solution, we are able to characterize the impact of liquidity on information acquisition behavior using numerical methods. We also consider quadratic transaction costs, as assumed by Garleanu and Pedersen (2013). Quadratic transaction costs capture the observation that the cost of immediacy is a convex function of trade size. Because such transaction costs are differentiable, we obtain a closed-form solution for the optimal information acquisition policy for the multi-asset case.

Although Van Nieuwerburgh and Veldkamp (2010) discuss a number of utility functions, our analysis focuses on the mean-variance utility function, which is one of the most widely used utility functions. Moreover, we consider both an additive learning technology and an entropy learning technology. Our key findings can be summarized as follows. First, as might be expected, we demonstrate that the liquidity of an asset affects the attractiveness of learning about it. As the trading costs of an asset increase, it becomes less attractive for investors to learn about it; this conclusion holds for various learning technologies and transaction costs functions. Second, when introducing a bundle which has lower transaction costs than trading all the constituent stocks, we find that only those investors with additive learning capacity diversify their information allocations. Although trading the bundle is cheaper than trading the constituent stocks, investors with entropy learning technology choose to learn about only one asset. Third, investors equipped with different learning capacities and starting allocations might choose to learn about different assets. Instead of having one asset that attracts the attention of all investors, investors with higher learning capacities typically prefer more liquid assets because such investors take more aggressive positions, while other investors are less sensitive to changes in transaction costs. Fourth, transaction costs not only affect the attractiveness of assets but also whether investors choose to learn about one or multiple assets. With mean-variance utility and without transaction costs, investors engage in specialized learning, meaning that they learn about a single asset. With transaction costs, the optimal information acquisition policy for investors with an additive learning technology might be combined learning, for both proportional and quadratic transaction costs. With quadratic transaction costs, investors with an entropy learning technology might also use generalized learning to achieve a higher utility meaning that they learn about multiple assets. Hence, introducing transaction costs not only affects the relative attractiveness of assets but also investors' learning behavior.

This paper is related to the literature on endogenous information acquisition in financial markets (e.g., Grossman and Stiglitz (1980); Verrecchia (1982); Van Nieuwerburgh and Veldkamp (2010); Van Nieuwerburgh and Veldkamp (2009)). Our analysis also fits into a larger branch of the literature that addresses portfolio choice with frictions such as illiquidity. Illiquidity is modeled as either the inability to trade continuously (see Longstaff (2001) and Longstaff (2009) for example) or using transaction costs. Because our model setting is static, we take the second approach. We view illiquidity as a type of explicit transaction cost that investors pay when rebalancing their portfolios. Many papers investigate the optimal portfolio choice for an agent facing transaction costs. A few examples include Dybvig (2005) for the two-period case, and Dumas and Luciano (1991), Liu (2004), Liu and Loewenstein (2002), Lo, Mamaysky, and Wang (2004), and Constantinides, Jackwerth, and Perrakis (2009) for the dynamic setting. Our main contribution is to combine both information acquisition and transaction costs in a single setup in order to investigate how liquidity and information acquisition interact.

The remainder of this paper is organized as follows. Section 4.2 introduces the framework of the paper, including preferences, the investment opportunity set and the investor's portfolio and information choice problem. Section 4.3 presents the results with proportional transaction costs in the two-asset case. Section 4.4 considers the case of multiple assets but with quadratic transaction costs. Concluding remarks are offered in Section 4.5.

4.2 Model Setup

We borrow the three-period setup of Van Nieuwerburgh and Veldkamp (2010). In the first period, investors choose the precision of signals about asset payoffs. In the second period, investors observe the signal and then make their portfolio choices. In the last period, payoffs are realized.

4.2.1 Financial Assets

In our model, the available investment universe consists of one riskless asset, two risky assets, and a bundle that is a linear combination of the two risky assets. The constant risk-free rate is denoted by r. The two risky assets, with prices $\boldsymbol{P} = [P_1, P_2]'$, provide payoffs (returns) that are denoted by $\boldsymbol{f} = [f_1, f_2]'$. Assuming that the weight of asset 1 in the bundle is b, the bundle has an excess return of $bf_1 + (1-b)f_2$. The excess returns are not public information to investors.

Without loss of generality, we assume that risky assets are independent in our model, which is the same assumption made by Van Nieuwerburgh and Veldkamp (2010). This is without loss of generality since if assets are correlated, we can form principal components such that the linear combination of the initial assets are independent. We can interpret the independent assets so formed as risk factors. Because the risk factors are linear combinations of the initial assets, investment in the risky factors can easily be translated into investment in the initial assets.

Trading these three assets involves proportional transaction costs. Let t_1, t_2, t_3 denote the corresponding proportional trading costs for trading asset 1, asset 2, and the bundle. The same transaction cost is incurred whether buying or selling assets. We assume that $t_3 \leq bt_1 + (1-b)t_2$, i.e., trading the bundle is cheaper than trading the constituent stocks. The bundle in our model can be interpreted as an index consisting of all the assets in the model (in our simple model, there are only 2 assets). It would be interesting to determine whether the convenience of trading an index product leads investors to learn about multiple assets.

4.2.2 Investors' Beliefs

In the first period, investors share a common prior belief regarding the payoffs of the risky assets in the third period, which is captured by a normal distribution $\boldsymbol{f} \sim (\boldsymbol{\mu}, \boldsymbol{\Sigma})$. Given the independence of $\boldsymbol{f}$, $\boldsymbol{\Sigma}$ is a diagonal matrix. In the first period, investors can acquire information regarding payoffs by obtaining unbiased but noisy signals of the payoff $\boldsymbol{\eta} = \boldsymbol{f} + \boldsymbol{e}_\eta$, where $\boldsymbol{e}_\eta \sim N(\boldsymbol{0}, \boldsymbol{\Sigma}_\eta)$. We assume that the signals are also independent, i.e. $\boldsymbol{\Sigma}_\eta$ is also a diagonal matrix. The smaller the diagonal elements of $\boldsymbol{\Sigma}_\eta$ are, the closer the signals are to actual payoffs. The precision of the signal $\boldsymbol{\eta}$ hinges on how investors allocate their learning capacity. Obviously, when investors do not want to learn about asset i, $\Sigma_\eta(i,i) = \infty$.

In the first period, investors make their information acquisition decision by choosing the precision matrix $\boldsymbol{\Sigma}_\eta$. In the second period, they observe the signal and update their beliefs about the payoff using Bayes' rule. We let E_2 denote the posterior expectation combining the prior belief and the signal. Let $\hat{\boldsymbol{\mu}}$ and $\hat{\boldsymbol{\Sigma}}$ denote the payoff's mean and variance conditional on the signal respectively. Using Bayes' rule, combining the noisy signals and the ex ante belief, we obtain

$$\hat{\boldsymbol{\mu}} = E_2[\boldsymbol{f} \mid \boldsymbol{\mu}, \boldsymbol{\eta}] = (\boldsymbol{\Sigma}^{-1} + \boldsymbol{\Sigma}_{\boldsymbol{\eta}}^{-1})^{-1}(\boldsymbol{\Sigma}^{-1}\boldsymbol{\mu} + \boldsymbol{\Sigma}_{\boldsymbol{\eta}}^{-1}\boldsymbol{\eta}) , \tag{4.1}$$
$$\hat{\boldsymbol{\Sigma}} = V_2[\boldsymbol{f} \mid \boldsymbol{\mu}, \boldsymbol{\eta}] = (\boldsymbol{\Sigma}^{-1} + \boldsymbol{\Sigma}_{\boldsymbol{\eta}}^{-1})^{-1} . \tag{4.2}$$

Based on these updated beliefs, investors then make their portfolio decision. Payoffs are realized in period 3.

4.2.3 Investors' Utility

Investors' preferences are captured by a utility function denoted by $u_2(W)$, where W denotes terminal wealth at time 3. Because investors can learn the payoff at time 2, we let u_1 denote another utility function to capture investors' preference for early resolution of uncertainty. Such a utility function is considered under the expectations of the prior belief, E_1. In period 1, an investor's utility is defined as

$$U_1 = E_1\left[u_1\left(E_2[u_2(W)]\right)\right] . \tag{4.3}$$

The curvature of u_1 determines whether investors prefer early or late resolution of uncertainty. Because expectation E_2 depends on the outcome of the signal, learning at time 2 makes $E_2[u_2(W)]$ a random variable. The randomness of the expectation gives higher utility to investors with convex u_1 and lower utility to investors with concave u_1. Hence, investors with convex u_1 prefer early resolution of uncertainty, while investors with concave u_1 prefer late resolution of uncertainty. When u_1 is a linear function, investors are indifferent as to when the information is disclosed.

Although many types of preferences have been discussed in the literature, we focus on one of the most commonly used types of utility. We assume that $u_1(x) = -\ln(-x)$ and $u_2(W) = -\exp(-\gamma W)$. Because payoffs are normally distributed, such a utility function is equivalent to a mean-variance utility function:

$$U_1 = E_1[\gamma E_2[W] - \frac{\gamma^2}{2} V_2[W]] , \tag{4.4}$$

where V_2 denotes the variance of terminal wealth conditional on the signal. Assume that investors are initially endowed with $\boldsymbol{\theta^0}$ of risky assets which have a total value W_0. Let $\boldsymbol{\theta}$ denote investors' holdings in period 2. Letting $TC(\cdot)$ denote the trading cost function, investors are subject to the following budget constraint:

$$W = W_0 r + \boldsymbol{\theta}'(\boldsymbol{f} - \boldsymbol{P}r) - TC(\boldsymbol{\theta} - \boldsymbol{\theta^0}) . \tag{4.5}$$

Substituting wealth back into the utility function, we obtain

$$U_1 = E_1\left[\gamma E_2[W_0 r + \boldsymbol{\theta}'(\boldsymbol{f} - \boldsymbol{P}r) - TC(\boldsymbol{\theta} - \boldsymbol{\theta^0})] - \frac{\gamma^2}{2} V_2[\boldsymbol{\theta}'(\boldsymbol{f} - \boldsymbol{P}r)]\right] . \tag{4.6}$$

Since r, γ and W_0 are not choice variables and are multiplicative constants, we drop these without changing the optimization problem. The time 1 utility simplifies to

$$U_1 = E_1\left[E_2[\boldsymbol{\theta}'(\boldsymbol{f} - \boldsymbol{P}r) - TC(\boldsymbol{\theta} - \boldsymbol{\theta^0})] - \frac{\gamma}{2} V_2[\boldsymbol{\theta}'(\boldsymbol{f} - \boldsymbol{P}r)]\right] . \tag{4.7}$$

To derive optimal information and portfolio choice, we must first solve for the optimal portfolio as a function of posterior beliefs and then choose the posterior belief variance that maximize the utility U_1.

4.2.4 Portfolio Choice with Given Beliefs

For a given belief about asset payoffs $(\hat{\boldsymbol{\mu}}, \hat{\boldsymbol{\Sigma}})$, investors maximize the following utility function in period 2:

$$U(\hat{\boldsymbol{\mu}}, \hat{\boldsymbol{\Sigma}}) = \boldsymbol{\theta}'(\hat{\boldsymbol{\mu}} - \boldsymbol{P}r) - TC(\boldsymbol{\theta} - \boldsymbol{\theta^0}) - \frac{\gamma}{2} \boldsymbol{\theta}' \hat{\boldsymbol{\Sigma}} \boldsymbol{\theta} . \tag{4.8}$$

Obviously, we have $U_1 = E_1(U)$. The portfolio after trading in period 2 is $\boldsymbol{\theta} = \boldsymbol{\theta}^0 + \boldsymbol{T}\boldsymbol{q}$, where $\boldsymbol{q} = [q_1, ..., q_6]$ denotes a vector indicating the selling/buying of risky asset 1, 2, and the bundle. We have $\boldsymbol{q} \geq 0$, i.e., when investors buy or sell i, $\boldsymbol{q}$ takes positive values in the corresponding position. The matrix $\boldsymbol{T}$ can be written as

$$\boldsymbol{T} = \begin{bmatrix} -1 & 1 & 0 & 0 & -b & b \\ 0 & 0 & -1 & 1 & b-1 & 1-b \end{bmatrix} . \tag{4.9}$$

In our model, transaction costs are proportional. Hence, $TC(\boldsymbol{\theta} - \boldsymbol{\theta}^0) = \boldsymbol{Cq}$, where

$$\boldsymbol{C} = [t_1, t_1, t_2, t_2, t_3, t_3]'. \tag{4.10}$$

We can write the utility U in terms of $\boldsymbol{q}$ as follows:

$$U(\theta) = (\boldsymbol{\theta}^0 + \boldsymbol{Tq})'(\hat{\boldsymbol{\mu}} - \boldsymbol{P}r) - \frac{\gamma}{2}(\boldsymbol{\theta}^0 + \boldsymbol{Tq})'\hat{\boldsymbol{\Sigma}}(\boldsymbol{\theta}^0 + \boldsymbol{Tq}) - \boldsymbol{C}'\boldsymbol{q} \, . \tag{4.11}$$

The Lagrangian function of the optimization problem is

$$\mathscr{L}(\boldsymbol{q}) = (\boldsymbol{\theta}^0 + \boldsymbol{Tq})'(\hat{\boldsymbol{\mu}} - \boldsymbol{P}r) - \frac{\gamma}{2}(\boldsymbol{\theta}^0 + \boldsymbol{Tq})'\hat{\boldsymbol{\Sigma}}(\boldsymbol{\theta}^0 + \boldsymbol{Tq}) - \boldsymbol{C}'\boldsymbol{q} + \boldsymbol{\lambda}'\boldsymbol{q} \, , \tag{4.12}$$

subject to

$$\boldsymbol{q} \geq 0 \, . \tag{4.13}$$

The complementary slackness condition is

$$\lambda_i q_i = 0, \quad \text{for } i = 1,..,6. \tag{4.14}$$

The first-order condition with respect to $\boldsymbol{q}$ is

$$\frac{\partial \mathscr{L}}{\partial \boldsymbol{q}} = \boldsymbol{T}'(\hat{\boldsymbol{\mu}} - \boldsymbol{P}r) - \gamma \boldsymbol{T}'\hat{\boldsymbol{\Sigma}}(\boldsymbol{\theta}^0 + \boldsymbol{Tq}) - \boldsymbol{C} + \boldsymbol{\lambda} \, . \tag{4.15}$$

The second-order derivative is

$$\frac{\partial^2 \mathscr{L}}{\partial \boldsymbol{q}^2} = -\gamma \boldsymbol{T}'\hat{\boldsymbol{\Sigma}}\boldsymbol{T} \, , \tag{4.16}$$

which is a negative semi-definite matrix. Therefore, the Lagrangian function is concave, and we have interior solutions. When comparing different candidates to search for the optimal solution, we can eliminate some obvious interior points that are not optimal. For the same target position, investors adopt the most cost-efficient way to trade. For example, when investors attempt to buy both assets, investors first buy the bundle and then fill the remaining position by trading the individual asset. In the end, we obtain the following 13 cases by eliminating trading that is not cost efficient:

(1) $q_1 > 0$; (2) $q_2 > 0$; (3) $q_3 > 0$; (4) $q_4 > 0$; (5) $q_5 > 0$; (6) $q_6 > 0$;
(7) $q_1 > 0, q_4 > 0$; (8) $q_2 > 0, q_3 > 0$; (9) $q_1 > 0, q_5 > 0$; (10) $q_2 > 0, q_6 > 0$;

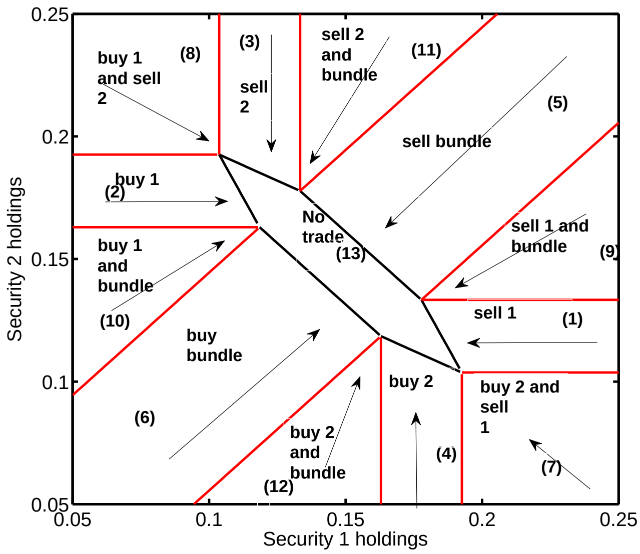

FIGURE 4.1: *Portfolio choice with proportional transaction costs.*
This figure displays investors' portfolio choice. The parameters are the following: risk aversion $\gamma = 5$, the expected return and volatility of both risky assets $\hat{\mu}_1 - rP_1 = \hat{\mu}_2 - rP_2 = 0.1$ and $\sqrt{\hat{\boldsymbol{\Sigma}}_{11}} = \sqrt{\hat{\boldsymbol{\Sigma}}_{22}} = 0.3$, respectively. Transaction costs of both risky assets are $t_1 = t_2 = 0.01$, while the transaction costs of the bundle are $t_3 = 0.005$. The number in parenthesis in each region corresponds to the case number.

(11) $q_3 > 0, q_5 > 0$; (12) $q_4 > 0, q_6 > 0$; (13) do nothing .
Due to the complexity of the computations, we leave the exact solutions to the 13 cases to Appendix 4.A.

Figure 4.1 illustrates these 13 cases as a function of the investor's initial allocation. Investors who begin in the no-trade region which is represented by the interior of the hexagon would simply retain their initial allocation. When there are trading costs, rebalancing all the way to the ideal portfolio is not profitable. When investors begin with an allocation outside of the no-trade region, they trade until they reach the target, i.e., the boundaries of the hexagon. There are 12 transaction regions. The arrows represent the transaction directions in these transaction regions. For example, in the buy 1 and sell 2 region (the quadrant on the top left), investors buy asset 1 and sell asset 2 to reach the target.

4.2.5 Learning Technology

Since the ex ante variance $\boldsymbol{\Sigma}$ is given, the posterior variance $\hat{\boldsymbol{\Sigma}}$ computed in equation (4.2) is not random. Every signal variance $\boldsymbol{\Sigma}_{\boldsymbol{\eta}}$ has a unique posterior belief variance $\hat{\boldsymbol{\Sigma}}$ associated with it. Hence, following Van Nieuwerburgh and Veldkamp (2010), we can define information choice as optimizing over posterior belief variance $\hat{\boldsymbol{\Sigma}}$ directly rather than over the signal variance $\boldsymbol{\Sigma}_{\boldsymbol{\eta}}$. The smaller the diagonal matrix elements of $\hat{\boldsymbol{\Sigma}}$ are, the more precise the posterior belief is.

How much information investors can acquire depends on their overall learning capacity, which we denote by K. We consider two types of learning technologies in our model, an additive learning technology and an entropy learning technology, which are also discussed in Van Nieuwerburgh and Veldkamp (2010).

Additive learning defines the learning capacity K as the sum of the differences between the inverse of the posterior and prior precision of each asset

$$\sum_{i=1}^{2} \left(\hat{\boldsymbol{\Sigma}}_{ii}^{-1} - \boldsymbol{\Sigma}_{ii}^{-1}\right) \leq K \, . \tag{4.17}$$

Entropy learning defines capacity K as the sum of the ratio of the precision, i.e.,

$$\sum_{i=1}^{2} \ln\left(\hat{\boldsymbol{\Sigma}}_{ii}^{-1} \boldsymbol{\Sigma}_{ii}\right) \leq K \, . \tag{4.18}$$

Furthermore, we also impose the non-forgetting constraint to ensure that investors cannot obtain more precise information about one asset by forgetting the prior information about another asset. Mathematically, the constraint can be written as

$$0 \leq \hat{\boldsymbol{\Sigma}}_{ii} \leq \boldsymbol{\Sigma}_{ii}, \text{ or equivalently, } \infty \geq \hat{\boldsymbol{\Sigma}}_{ii}^{-1} \geq \boldsymbol{\Sigma}_{ii}^{-1} \text{ for } i = 1, 2 \, . \tag{4.19}$$

4.2.6 The Information Choice Problem

Armed with the optimal portfolio $\boldsymbol{\theta}^*$ for a given belief $\left(\hat{\boldsymbol{\mu}}, \hat{\boldsymbol{\Sigma}}\right)$, we can derive investors' optimal information choice by means of the following maximization problem:

$$\max_{\hat{\boldsymbol{\Sigma}}} E_1 \left[\left(\boldsymbol{\theta}^* \left(\hat{\boldsymbol{\mu}}, \hat{\boldsymbol{\Sigma}}\right)\right)^T \left(\hat{\boldsymbol{\mu}} - \boldsymbol{P} r\right) - TC\left(\boldsymbol{\theta}^* \left(\hat{\boldsymbol{\mu}}, \hat{\boldsymbol{\Sigma}}\right) - \boldsymbol{\theta}^{\mathbf{0}}\right) - \frac{\gamma}{2} \left(\boldsymbol{\theta}^* \left(\hat{\boldsymbol{\mu}}, \hat{\boldsymbol{\Sigma}}\right)\right)^T \hat{\boldsymbol{\Sigma}} \left(\boldsymbol{\theta}^* \left(\hat{\boldsymbol{\mu}}, \hat{\boldsymbol{\Sigma}}\right)\right) \right] \, , \tag{4.20}$$

subject to the learning constraint

$$\sum_{i=1}^{2} \ln\left(\hat{\boldsymbol{\Sigma}}_{ii}^{-1} \boldsymbol{\Sigma}_{ii}\right) \leq K \quad \text{or} \quad \sum_{i=1}^{2} \left(\hat{\boldsymbol{\Sigma}}_{ii}^{-1} - \boldsymbol{\Sigma}_{ii}^{-1}\right) \leq K \, , \tag{4.21}$$

and the non-forgetting constraint

$$\infty \geq \hat{\boldsymbol{\Sigma}}_{ii}^{-1} \geq \boldsymbol{\Sigma}_{ii}^{-1}, \text{for } i = 1, 2 \, . \tag{4.22}$$

The expectation E_1 defined in (4.20) is taken over the posterior mean $\hat{\boldsymbol{\mu}}$. $\hat{\boldsymbol{\mu}}$ depends on the realization of signals and is a random variable in the first period. It follows a normal distribution with the following mean and variance

$$E_1(\hat{\boldsymbol{\mu}}) = (\boldsymbol{\Sigma}^{-1} + \boldsymbol{\Sigma}_{\boldsymbol{\eta}}^{-1})^{-1}(\boldsymbol{\Sigma}^{-1}\boldsymbol{\mu} + \boldsymbol{\Sigma}_{\boldsymbol{\eta}}^{-1}E_1(\boldsymbol{\eta})) = \boldsymbol{\mu}\,, \tag{4.23}$$

$$\mathrm{V}_1(\hat{\boldsymbol{\mu}}) = \hat{\boldsymbol{\Sigma}}^{-1}\boldsymbol{\Sigma}_{\boldsymbol{\eta}}^{-1}V_1(\boldsymbol{\eta})\boldsymbol{\Sigma}_{\boldsymbol{\eta}}^{-1}\hat{\boldsymbol{\Sigma}} = \hat{\boldsymbol{\Sigma}}^{-1}\boldsymbol{\Sigma}_{\boldsymbol{\eta}}^{-1}(\boldsymbol{\Sigma} + \boldsymbol{\Sigma}_{\boldsymbol{\eta}})\boldsymbol{\Sigma}_{\boldsymbol{\eta}}^{-1}\hat{\boldsymbol{\Sigma}} = \boldsymbol{\Sigma} - \hat{\boldsymbol{\Sigma}}\,. \tag{4.24}$$

The last equality in (4.24) follows from the fact that $\boldsymbol{\Sigma}$ and $\boldsymbol{\Sigma}_{\boldsymbol{\eta}}$ are diagonal matrices.

Unlike Van Nieuwerburgh and Veldkamp (2010), in which optimal portfolio holdings are a smooth function of posterior beliefs, optimal portfolio holdings in the case of proportional transaction costs are a non-smooth function of the posterior beliefs and cannot be differentiated. Therefore, a closed-form solution is not possible.

For a given $\hat{\boldsymbol{\Sigma}}$, the expectation E_1 in equation (4.20) can be computed as the sum of the integrals of function U over 13 different regions corresponding to the 12 trading regions and the no-trade region. The precise functional form of U differs in the 13 regions and is presented in Appendix 4.B together with the boundaries of the 13 regions. Having the integrand U in closed-form, numerical integration allows us to find the optimal choice of posterior belief variance much faster than using simulation methods.

4.3 Two Risky Assets and Proportional Transaction Costs

Van Nieuwerburgh and Veldkamp (2010) consider a case involving mean-variance utility and entropy-learning capacity and conclude that investors spend all their capacity in learning about the asset with the highest squared Sharpe ratio. In our setting with transaction costs, the results are more complicated. Initial allocations, transaction costs, and the Sharpe ratio all play a role in determining investors' optimal information acquisition. Because we consider two types of learning technology that generate different results, we report the results in two subsections.

4.3.1 Results for Additive Learning Technology

We first consider the additive learning technology and observe the following properties.

Proposition 4 *When a risky asset's trading costs increase, learning about it becomes less attractive to investors.*

When an asset's transaction costs increase, the expected net return of trading the asset decreases which makes learning about it less profitable. This effect is illustrated in Panel A of Figure 4.2. In the case considered in the figure, there are two risky assets and the bundle is not available for trading. Assets 1 and 2 have the same volatility, but asset 1 has a higher expected return. As the transaction costs of asset 1 increase, its attractiveness decreases. Once the

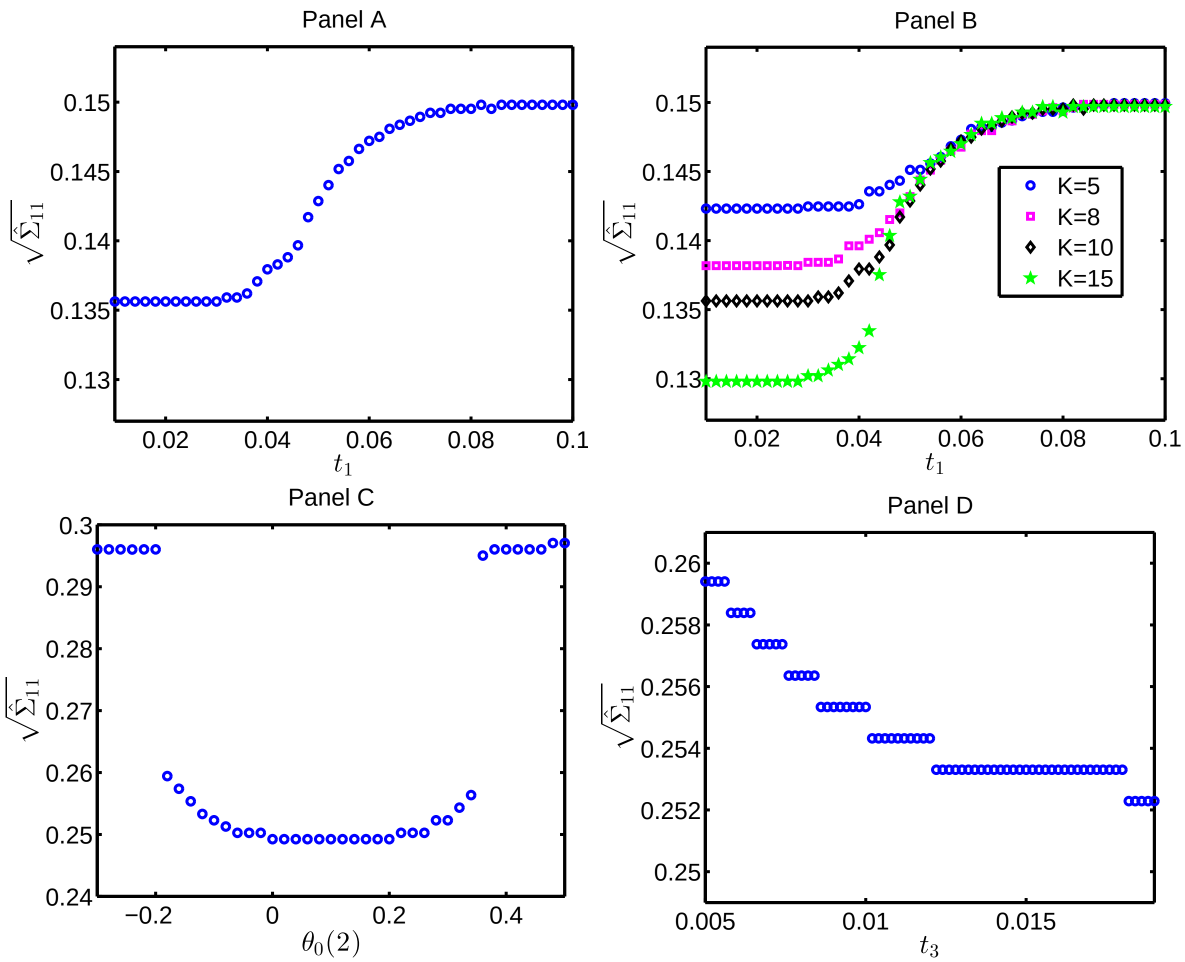

FIGURE 4.2: *Information acquisition with proportional transaction costs and an additive learning technology.*

This figure shows the optimal choice of learning about asset 1 measured in terms of the posterior volatility $\sqrt{\hat{\mathbf{\Sigma}}_{11}}$. Panel A shows the case in which the transaction costs of asset 1 vary. Panel B extends Panel A by allowing for various learning capacities. Panel C plots the case in which the initial endowment of asset 2 changes. The parameters for Panel A are $\mu_1 - rP_1 = 0.1$, $\mu_2 - rP_2 = 0.05$, $\sqrt{\mathbf{\Sigma}_{11}} = 0.15$, $\sqrt{\mathbf{\Sigma}_{22}} = 0.15$, $t_2 = 0.01$, $K = 10$, $\theta^0(1) = 0$, $\theta^0(2) = 0$, and $\gamma = 5$. In Panel B, the parameter values are the same as those in Panel A, and we allow learning capacity to vary from $K = 5$ to 8, 10, and 15. The parameters for Panel C are $\mu_1 - rP_1 = 0.1$, $\mu_2 - rP_2 = 0.05$, $\sqrt{\mathbf{\Sigma}_{11}} = 0.3$, $\sqrt{\mathbf{\Sigma}_{22}} = 0.3$, $t_1 = 0.02$, $t_2 = 0.02$, $K = 10$, $\theta^0(1) = 0.3$ and $\gamma = 5$. In Panel D, the parameters are the same as those in Panel C and $\theta^0(2) = 0.3$. Moreover, we allow for trading the bundle in this figure, and we set the weight $b = 0.5$.

cost exceeds a certain level, investors start learning about both assets. When asset 1 becomes very expensive to trade with $t_1 = 0.08$, investors choose not to learn about asset 1.[1] Indeed, in the setting with trading costs, expected return and volatility are not sufficient to measure the asset's attractiveness. Investors take transaction costs into account when comparing assets.

Proposition 5 *Investors with various learning capacities might choose to learn about different assets.*

Although increased transaction costs of one asset generally lead investors to learn about other assets, the magnitude of the impact tends to be larger for those investors with higher learning capacities. This relationship is illustrated in Panel B of Figure 4.2 , in which we allow learning capacity to vary. Investors with low learning capacity, $K = 5$, continue to learn exclusively about asset 1 when t_1 increases from 0.01 to 0.04. Investors with higher learning capacity $K = 10$ begin to generalize their learning at the point at which $t_1 = 0.028$. It is likely that investors with higher skills take more aggressive positions than less skilled investors; therefore, their information acquisition is more sensitive to changes in trading costs. Consequently, investors with different learning capacity might choose to learn about different assets. These results differ from the case with no transaction costs, in which all investors choose to learn about only one asset.

Proposition 6 *When investors' initial asset allocations vary, investors choose to learn about different assets.*

In a setting with transaction costs, the starting allocation has an impact on information acquisition. Consider the example shown in Panel C of Figure 4.2, in which both assets have the same transaction costs and ex ante volatility, but asset 1 has a higher expected return. When the initial holdings of asset 2 are extreme, either below -20% or above 35%, investors only learn about asset 2. However, when the initial holdings of asset 2 are less extreme, between -20% and 0% and between 20% and 35%, investors learn about both assets. Finally, when the initial holdings of asset 2 are between 0% and 20%, investors only learn about asset 1. The optimal precision of asset 1 is a U-shaped curve of the initial holdings of asset 2. The reason for such a learning curve is simple: when making their information acquisition decision, investors also consider transaction costs. Therefore, investors choose to learn about those assets that they expect to trade. When the initial allocation of asset 2 is extreme, investors expect to trade asset 2 and spend all their capacity in learning about this asset's payoff. When the initial allocation of asset 2 is less extreme, asset 1, which has a higher Sharpe ratio, becomes more attractive for investors to learn about. It is worth noting that without transaction costs no matter where investors start, they always make the same learning decision. However, when there are transaction costs, investors might choose to learn about different assets depending on their starting allocations.

[1] The optimal information choice with additive learning and mean-variance utility but without transaction costs is not discussed by Van Nieuwerburgh and Veldkamp (2010). We solve this case in Section 4.4.1.

Proposition 7 *When the transaction costs of the bundle decrease, combined learning becomes more attractive.*

This property is illustrated in Panel D of Figure 4.2 where we allow for the trading of the bundle. In our example, assets 1 and 2 are identical except that asset 1 has a higher expected return. Both assets have equal weights in the bundle, i.e. $b = 0.5$. When the bundle has the same trading costs as individual assets, investors choose to learn about asset 1 only. When the bundle's transaction costs t_3 decrease, investors spend less capacity on learning about asset 1 and begin learning about both assets. Effectively, the ability to trade the bundle at low cost makes changing the allocation to asset 2 cheaper, and therefore acquiring information about it more attractive.

4.3.2 Results for Entropy Learning Technology

As discussed in Van Nieuwerburgh and Veldkamp (2010), one drawback of the additive learning technology is that it is not scale neutral in the sense that share splits or reverse splits change the feasible information set. Therefore, we also consider entropy learning, which is scale-neutral. Although many results derived from additive learning still hold, there are certain differences, and we report the results in this section.

Proposition 8 *When a risky asset's trading costs increase, learning about it becomes less attractive to investors.*

This property is identical to the case of additive learning technology. In addition to the Sharpe ratio, transaction costs play a role in determining optimal information acquisition. Moreover, as shown in Panel A of Figure 4.3 , with the entropy learning technology, investors choose to learn about a single asset, i.e. use specialized learning. This observation is consistent with Van Nieuwerburgh and Veldkamp (2010), who also find that the additive learning technology favors generalized learning and entropy learning technology favors specialized learning. The reason is that the second-order condition for the learning constraint with respect to Σ_{ii}^{-1} is zero for additive learning, whereas it is negative for entropy learning. Hence, the second-order condition of the Lagrangian with respect to Σ_{ii}^{-1} is more likely to be positive when the constraint is entropy rather than additive learning because the learning constraint enters negatively into the Lagrangian function.

Proposition 9 *Investors with various learning capacities might choose to learn about different assets.*

This property is the same as for additive learning technology. However, since it is generally optimal to learn about a single asset, Panel B of Figure 4.3 shows that investors choose to learn about either asset 1 or asset 2. The lower capacity, the higher the level of t_1 required to induce investors to learn about asset 2.

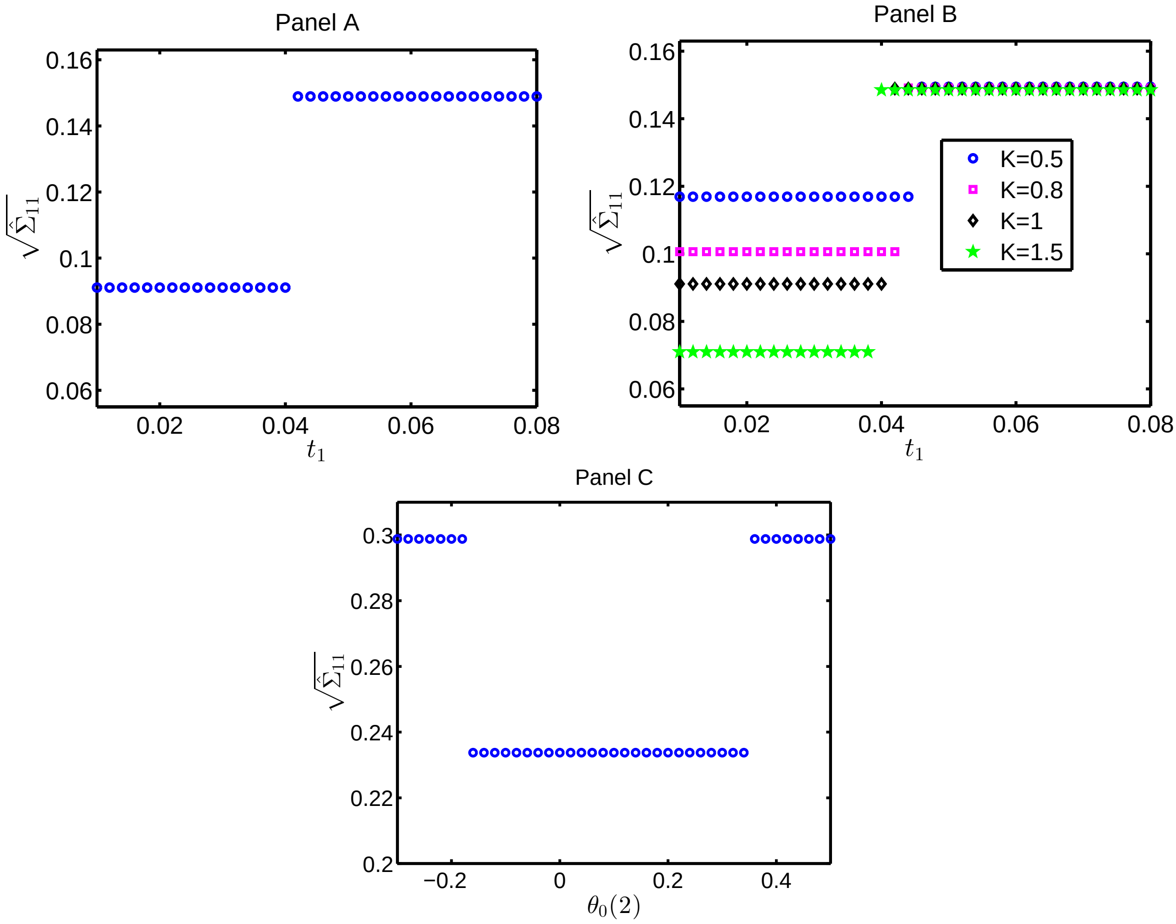

FIGURE 4.3: *Information acquisition with proportional transaction costs and an entropy learning technology.*

This figure shows the optimal choice of learning about asset 1 measured in terms of the posterior volatility $\sqrt{\hat{\mathbf{\Sigma}}_{11}}$. Panel A shows the case in which the transaction costs of asset 1 vary. Panel B extends Panel A by allowing for various learning capacities. Panel C plots the case in which the initial allocation of asset 2 changes. The parameters for Panel A are $\mu_1 - rP_1 = 0.1$, $\mu_2 - rP_2 = 0.05$, $\sqrt{\mathbf{\Sigma}_{11}} = 0.15$, $\sqrt{\mathbf{\Sigma}_{22}} = 0.15$, $t_2 = 0.01$, $K = 1$, $\theta^0(1) = 0$, $\theta^0(2) = 0$, and $\gamma = 5$. In Panel B, the parameter values are the same as those in Panel A, and we allow learning capacity to vary from $K = 0.5$ to 0.8, 1, and 1.5. The parameters for Panel C are $\mu_1 - rP_1 = 0.07$, $\mu_2 - rP_2 = 0.05$, $\sqrt{\mathbf{\Sigma}_{11}} = 0.3$, $\sqrt{\mathbf{\Sigma}_{22}} = 0.3$, $t_1 = 0.02$, $t_2 = 0.02$, $K = 0.5$, $\theta^0(1) = 0.3$ and $\gamma = 5$.

Proposition 10 *When investors' initial asset allocations vary, investors choose to learn about different assets.*

This property is identical to the one we observe for additive learning technology. Extreme initial allocations in a given asset induce investors to learn about that asset, while balanced initial allocations induce investors to learn about assets that have higher expected returns. Therefore, this property is robust to different definitions of learning technology.

Proposition 11 *No matter how low the transaction costs of the bundle are, combined learning never occurs.*

Although combined learning might be optimal for investors with additive learning technology, this is not the case for entropy learning technology. With the possibility to trade the bundle, it appears that it is never optimal for investors with entropy learning technology to learn about both assets, no matter how cheap trading the bundle is.[2]

4.4 Multiple Assets and Quadratic Transaction Costs

With proportional transaction costs, the no-trade region becomes complicated when the number of risky assets exceeds two. Although the no-trade region can still be characterized in closed form, solving equation (4.20) requires integrating a complicated piecewise quadratic function with a complex domain. Computing the objective function in each region is time-consuming.[3] Nonetheless, the qualitative results we obtain for two assets should remain the same.

In this section, we consider the case of an arbitrary number of assets N in a setting with quadratic transaction costs. More specifically, we assume that the transaction costs incurred when rebalancing the portfolio by $\boldsymbol{\theta} - \boldsymbol{\theta^0}$ are $\frac{1}{2}\left(\boldsymbol{\theta} - \boldsymbol{\theta^0}\right)' \boldsymbol{\Lambda} \left(\boldsymbol{\theta} - \boldsymbol{\theta^0}\right)$ and that $\boldsymbol{\Lambda} \geq \mathbf{0}$ is a diagonal matrix such that there are no correlated transaction costs.

To keep the analysis tractable, a bundle that is cheaper to trade is not available. Due to its differentiability, the assumption of quadratic transaction costs makes the computation of the optimal portfolio rather straightforward and allows us to obtain a closed-form solution for optimal information acquisition.

The use of quadratic transaction costs can be justified because the cost of immediacy is usually a convex function of the trade size. A trader who requires immediate execution of a large

[2] We did not manage to prove that investors with entropy learning technology never learn about multiple assets. The result is based on 10,000 observations of different parameter values selected randomly.

[3] We can use simulations to estimate the integral; however, even for the case of two risky assets, at least 10,000 simulations are required to obtain results with 1% accuracy. Computing the utility function with multiple assets requires more intensive computations.

order may have only the first portion filled at the best bid or ask quote. The second portion is filled at the second-best bid or ask price in the order book, and so on. Thus, the cost of immediacy increases more than in proportion to the trade size, particularly for large trades. Such convexity can be captured by a quadratic transaction cost function.

To solve the information and portfolio choice problems we first determine the portfolio for a given belief $(\hat{\boldsymbol{\mu}}, \hat{\boldsymbol{\Sigma}})$ in the second period. With quadratic transaction costs, the utility function U, defined in equation (4.8), becomes

$$U = \boldsymbol{\theta}' \left(\hat{\boldsymbol{\mu}} - r\boldsymbol{P}\right) - \frac{\gamma}{2}\boldsymbol{\theta}'\hat{\boldsymbol{\Sigma}}\boldsymbol{\theta} - \frac{1}{2}\left(\boldsymbol{\theta} - \boldsymbol{\theta}^{\mathbf{0}}\right)' \boldsymbol{\Lambda} \left(\boldsymbol{\theta} - \boldsymbol{\theta}^{\mathbf{0}}\right) . \tag{4.25}$$

The first-order condition is:

$$\left(\hat{\boldsymbol{\mu}} - r\boldsymbol{P}\right) - \gamma\hat{\boldsymbol{\Sigma}}\boldsymbol{\theta} - \boldsymbol{\Lambda}\left(\boldsymbol{\theta} - \boldsymbol{\theta}^{\mathbf{0}}\right) = \mathbf{0} . \tag{4.26}$$

Solving gives the optimal portfolio

$$\boldsymbol{\theta}^* = \left(\gamma\hat{\boldsymbol{\Sigma}} + \boldsymbol{\Lambda}\right)^{-1} \left(\hat{\boldsymbol{\mu}} - r\boldsymbol{P} + \boldsymbol{\Lambda}\boldsymbol{\theta}^{\mathbf{0}}\right) = \boldsymbol{A}^{-1}\left(\hat{\boldsymbol{\mu}} - r\boldsymbol{P} + \boldsymbol{\Lambda}\boldsymbol{\theta}^{\mathbf{0}}\right) , \tag{4.27}$$

where $\boldsymbol{A} \equiv \gamma\hat{\boldsymbol{\Sigma}} + \boldsymbol{\Lambda}$.

Substituting the solution back into the objective function (4.25), simplifying and taking the expectation of $\hat{\boldsymbol{\mu}}$, the first period utility function U_1 as defined in (4.7) becomes

$$U_1 = \frac{1}{2}\mathrm{Tr}\left(\boldsymbol{A}^{-1}(\boldsymbol{\Sigma} - \hat{\boldsymbol{\Sigma}})\right) + \frac{1}{2}\left(\boldsymbol{\mu} - r\boldsymbol{P}\right)' \boldsymbol{A}^{-1}\left(\boldsymbol{\mu} - r\boldsymbol{P}\right) + (\boldsymbol{\theta}^{\mathbf{0}})'\boldsymbol{\Lambda}'\boldsymbol{A}^{-1}\left(\boldsymbol{\mu} - r\boldsymbol{P}\right) + \frac{1}{2}(\boldsymbol{\theta}^{\mathbf{0}})'(\boldsymbol{\Lambda}\boldsymbol{A}^{-1}\boldsymbol{\Lambda} - \boldsymbol{\Lambda})\boldsymbol{\theta}^{\mathbf{0}} . \tag{4.28}$$

Denote $a_i = \frac{1}{2\gamma}\Lambda_{ii} + \frac{1}{2}\left((\mu_i - rP_i) + \theta_i^0\Lambda_{ii}\right)^2 + \frac{1}{2}\Sigma_{ii}$. Obviously, $a_i > 0$ for all i. Appendix 4.C shows that the information acquisition optimization problem (ignoring some constant terms) reduces to

$$\max_{\hat{\boldsymbol{\Sigma}}} \sum_{i=1}^{n} \frac{a_i}{\gamma\hat{\Sigma}_{ii} + \Lambda_{ii}} , \tag{4.29}$$

subject to

$$\sum_{i=1}^{n} \ln\left(\hat{\boldsymbol{\Sigma}}_{ii}^{-1}\boldsymbol{\Sigma}_{ii}\right) \leq K \quad \text{or} \quad \sum_{i=1}^{n}\left(\hat{\boldsymbol{\Sigma}}_{ii}^{-1} - \boldsymbol{\Sigma}_{ii}^{-1}\right) \leq K , \tag{4.30}$$

$$\infty \geq \hat{\boldsymbol{\Sigma}}_{ii}^{-1} \geq \boldsymbol{\Sigma}_{ii}^{-1}, \text{for } \ i = 1, 2, ..., n . \tag{4.31}$$

4.4.1 Results for Additive Learning Technology

In this subsection, we examine the case with additive learning technology. Denoting $k_i = \hat{\mathbf{\Sigma}}_{ii}^{-1}$, the Lagrangian in this case is given by

$$\mathscr{L} = \sum_{i=1}^{n} \frac{a_i}{\frac{\gamma}{k_i} + \Lambda_{ii}} + \lambda_0 \left(K - \sum_{i=1}^{n} \left(k_i - \mathbf{\Sigma}_{ii}^{-1} \right) \right) + \sum_{i=1}^{n} \lambda_i \left(k_i - \mathbf{\Sigma}_{ii}^{-1} \right) . \tag{4.32}$$

The first and second-order derivatives with respect to k_i are

$$\frac{\partial \mathscr{L}}{\partial k_i} = \frac{a_i \gamma}{\left(\gamma + \Lambda_{ii} k_i \right)^2} - \lambda_0 + \lambda_i \tag{4.33}$$

and

$$\frac{\partial^2 \mathscr{L}}{\partial k_i^2} = - \frac{2 a_i \gamma}{\left(\gamma + \Lambda_{ii} k_i \right)^3} \Lambda_{ii} . \tag{4.34}$$

Since the Hessian of $\mathscr{L}$ is a diagonal matrix, it is negative (positive) definite if and only if all its diagonal elements are negative (positive). Because the second order derivative of $\mathscr{L}$ is negative for all i, the objective function is concave. For the complementary slackness condition $\lambda_i \left(k_i - \Sigma_{ii}^{-1} \right) = 0$ to hold, whenever $k_i > \mathbf{\Sigma}_{ii}^{-1}$, we have $\lambda_i = 0$. Defining a function $h_i(k_i) \equiv \frac{a_i \gamma}{(\gamma + \Lambda_{ii} k_i)^2}$, the optimal information choice is $k_i^* = \max(\mathbf{\Sigma}_{ii}^{-1}, h_i^{-1}(\lambda_0))$, where λ_0 solves $K = \sum_{i=1}^{n} \max(\mathbf{\Sigma}_{ii}^{-1}, h_i^{-1}(\lambda_0))$.

For quadratic transaction costs with additive learning technology, we observe properties similar to those observed in Section 4.3.1. Figure 4.4 illustrates a simple case with only two risky assets in which both assets have the same volatility but asset 1 has a higher excess return. Panel A of Figure 4.4 shows that when the transaction costs of both assets are the same ($\Lambda_{11} = \Lambda_{22} = 0.01$), investors choose to learn about asset 1 only. As the transaction costs of asset 1 increase, investors also begin to learn about asset 2. If transaction costs exceed 0.043, investors specialize in learning only about asset 2. This result is similar to that in the case with proportional transaction costs. In Panel B, we compare the information choices of investors with various learning capacities. Panel B shows that investors with a higher learning capacity switch earlier to combined learning as asset 1's transaction costs Λ_{11} increase. Consistent with the case of proportional trading costs, more skilled investors are more sensitive to trading costs. Panel C shows that as their starting allocations to asset 2 vary, investors choose different assets to learn about. The asset they learn about is usually the one that they expect to trade more. An extreme position in asset 2 induces investors to spend more capacity in learning about this asset. Compared with the case with proportional transaction costs, we need fairly high transaction costs to obtain the results in Panel C because for trading sizes below 1, using the same coefficient ($t_1 = \Lambda_{11}$ and $t_2 = \Lambda_{22}$) gives much lower transaction costs

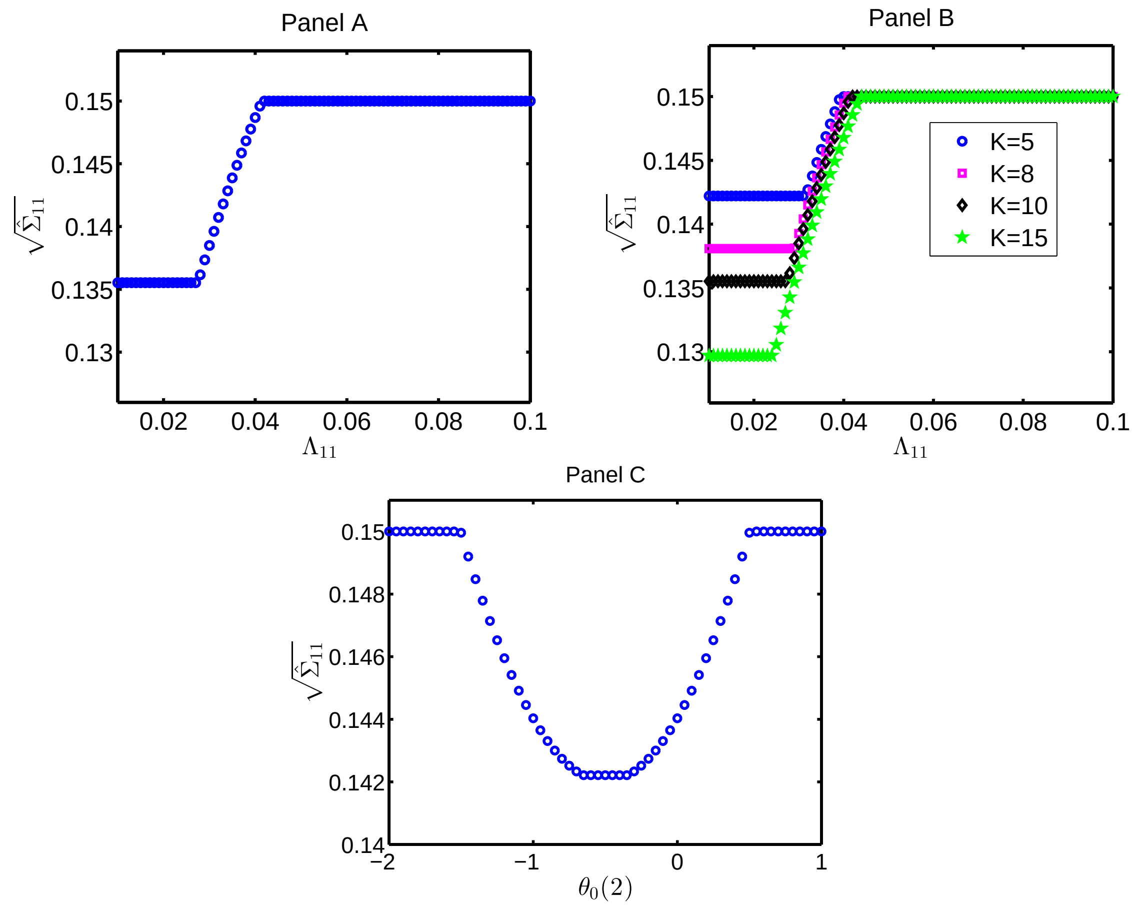

FIGURE 4.4: *Information acquisition with quadratic transaction costs and an additive learning technology.*

This figure shows the optimal choice of learning about asset 1 measured in terms of posterior volatility $\sqrt{\hat{\mathbf{\Sigma}}_{11}}$. Panel A shows the case in which asset 1's transaction costs vary. Panel B extends Panel A by allowing for various learning capacities. Panel C plots the case in which the initial endowment of asset 2 changes. The parameters for Panel A are $\mu_1 - rP_1 = 0.1$, $\mu_2 - rP_2 = 0.05$, $\sqrt{\mathbf{\Sigma}_{11}} = 0.15$, $\sqrt{\mathbf{\Sigma}_{22}} = 0.15$, $\Lambda_{22} = 0.01$, $K = 10$, $\theta^0(1) = 0$, $\theta^0(2) = 0$, and $\gamma = 5$. In Panel B, the parameter values are the same as those in Panel A, and we allow learning capacity to vary from $K = 5$ to 8, 10 and 15. The parameters for Panel C are $\mu_1 - rP_1 = 0.07$, $\mu_2 - rP_2 = 0.05$, $\sqrt{\mathbf{\Sigma}_{11}} = 0.15$, $\sqrt{\mathbf{\Sigma}_{22}} = 0.15$, $\Lambda_{11} = 0.1$, $\Lambda_{22} = 0.1$, $K = 10$, $\theta^0(1) = 0.3$ and $\gamma = 5$.

for the quadratic function than for proportional transaction costs.

In order to compare these results with those in the case with no transaction costs, we must solve the maximization problem when $\mathbf{\Lambda} = \mathbf{0}$. Since this case is not discussed in Van Nieuwerburgh and Veldkamp (2010), we briefly sketch the solution. In the special case in which $\mathbf{\Lambda} = \mathbf{0}$, the objective function is a weighted sum of posterior beliefs $\sum_{i=1}^{n} \frac{1}{\gamma}(\frac{1}{2}(\mu_i - rP_i)^2 + \frac{1}{2}\Sigma_{ii})k_i$. The learning capacity constraint bounds the sum of the precisions from above, whereas the non-forgetting condition bounds each k_i from below. The maximum of the Lagrangian function is reached at a corner. Specifically, investors choose to learn about the asset for which the expression $(\mu_i - rP_i)^2 + \Sigma_{ii}$ is highest. By contrast, with both quadratic and proportional transaction costs, investors choose to learn about multiple assets. Thus, transaction costs not only change the relative attractiveness of the risky assets but also have an impact on whether investors choose to engage in specialized or generalized learning. This shows that in the presence of transaction costs, the utility function and the learning technology alone do not fully determine whether investors engage in specialized or generalized learning.

4.4.2 Results for Entropy Learning Technology

We now investigate the solution for the entropy learning technology. Letting $K_i \equiv \ln\left(\frac{\Sigma_{ii}}{\hat{\Sigma}_{ii}}\right)$, we have $\hat{\Sigma}_{ii} = \frac{\Sigma_{ii}}{e^{K_i}}$ and the entropy learning technology constraint can be written in terms of K_i as $\Sigma_{i=1}^{n} K_i \leq K$. The non-forgetting constraint can be written as $K_i \geq 0$. The Lagrangian function can then be written in terms of K_i as follows:

$$\mathscr{L} = \sum_{i=1}^{n} \frac{a_i}{\frac{\gamma\Sigma_{ii}}{e^{K_i}} + \Lambda_{ii}} + \lambda_0\left(K - \Sigma_{i=1}^{n} K_i\right) + \sum_{i=1}^{n} \lambda_i K_i \,. \tag{4.35}$$

The first and second-order derivatives of $\mathscr{L}$ with respect to K_i are

$$\frac{\partial \mathscr{L}}{\partial K_i} = \frac{a_i \gamma \Sigma_{ii}}{(\frac{\gamma\Sigma_{ii}}{e^{K_i}} + \Lambda_{ii})^2 e^{K_i}} - \lambda_0 + \lambda_i \tag{4.36}$$

and

$$\frac{\partial^2 \mathscr{L}}{\partial K_i^2} = \frac{a_i \gamma \Sigma_{ii}}{(\frac{\gamma\Sigma_{ii}}{e^{K_i}} + \Lambda_{ii})^3 e^{2K_i}} \left(\gamma\Sigma_{ii} - \Lambda_{ii} e^{K_i}\right) \,. \tag{4.37}$$

Again, since the Hessian of $\mathscr{L}$ is a diagonal matrix, it is negative (positive) definite if and only if all its diagonal elements are negative (positive). Notably, the sign of the diagonal elements of the Hessian matrix hinges on the values of the parameters. It is obvious that the Hessian is positive definite if and only if $\gamma\Sigma_{ii} - \Lambda_{ii} e^{K_i} > 0$ for all i. If the opposite holds true for all i, the Hessian is negative definite. Because K_i is bounded by 0 and K, it follows directly that a sufficient condition for a positive definite Hessian matrix is that $K < \ln\left(\frac{\gamma\Sigma_{ii}}{\Lambda_{ii}}\right)$ holds for all i, which is the case when assets are very liquid (low Λ_{ii}), investors' prior uncertainty about the expected payoff Σ_{ii} is large, investors have low learning capacity K and their risk aversion γ is high. At the other extreme, when $\gamma\Sigma_{ii} - \Lambda_{ii}e^0 < 0$ for all i, the Hessian matrix is negative definite. The Hessian matrix is negative definite only when assets are highly illiquid and investors have accurate prior beliefs and low risk aversion.

We now investigate the properties of the optimal solution in these two cases. When the Hessian matrix is positive definite, or (in other words) the Lagrangian function is convex, the optimal solution is a corner solution in which investors spend all their capacity in learning about a single asset. The asset j about which investors choose to acquire information is the one for which the expression $-\frac{a_j}{\gamma\Sigma_{jj}+\Lambda_{jj}} + \frac{a_j}{\frac{\gamma\Sigma_{jj}}{e^K}+\Lambda_{jj}}$ is highest.

An example with two assets of a situation giving rise to specialized learning is shown in Figure 4.5. Overall, the results are very similar to those for the case of proportional transaction costs reported in Figure 4.3. Specifically, trading costs reduce the attractiveness of learning about a given asset (Panel A), overall learning capacity affects which assets investors choose to learn about (Panel B), as does their initial allocation (Panel C).

By contrast, when the Hessian matrix is negative definite, the Lagrangian function is concave, and we have interior solutions. For the complementary slackness condition $\lambda_i K_i = 0$ to hold, we must have $\lambda_i = 0$ when $K_i > 0$. Defining $g_i(K_i) = \frac{a_i\gamma\Sigma_{ii}}{(\frac{\gamma\Sigma_{ii}}{e^{K_i}}+\Lambda_{ii})^2 e^{K_i}}$, the optimal information choice is $K_i^* = \max(0, g_i^{-1}(\lambda_0))$, where λ_0 solves $K = \sum_{i=1}^n \max(0, g_i^{-1}(\lambda_0))$.

An example of a situation giving rise to generalized learning is shown in Figure 4.6. In this example, to ensure the concavity of the Lagrangian function, we use very high transaction costs and fairly low volatility, specifically $\Lambda_{11} = \Lambda_{22} = 0.1$ and $\sqrt{\mathbf{\Sigma}_{11}} = \sqrt{\mathbf{\Sigma}_{22}} = 0.1$. As was the case with entropy learning technology and proportional transaction costs, we observe the following: as trading costs increase, learning about the asset with the highest Sharpe ratio becomes less attractive to investors (Panel A). Investors with various learning capacities might choose to learn about different assets (Panel B). In addition, various starting allocations lead to different information choices (Panel C). In contrast to the case with proportional trading costs and to the case with no transaction costs, investors learn about multiple assets.

Our results show that for the entropy learning technology with quadratic transaction costs, whether specialized or generalized learning arises depends on the parameter values. A key driver of whether one or the other arises is how the convexity of the objective function in signal precision compares with the convexity of the learning capacity constraint. Because transaction costs affect the curvature of the objective function, it follows naturally that investors might switch from specialized to generalized learning when transaction costs are introduced.

Table 4.1 summarizes investors' information acquisition for all the cases considered in the paper.

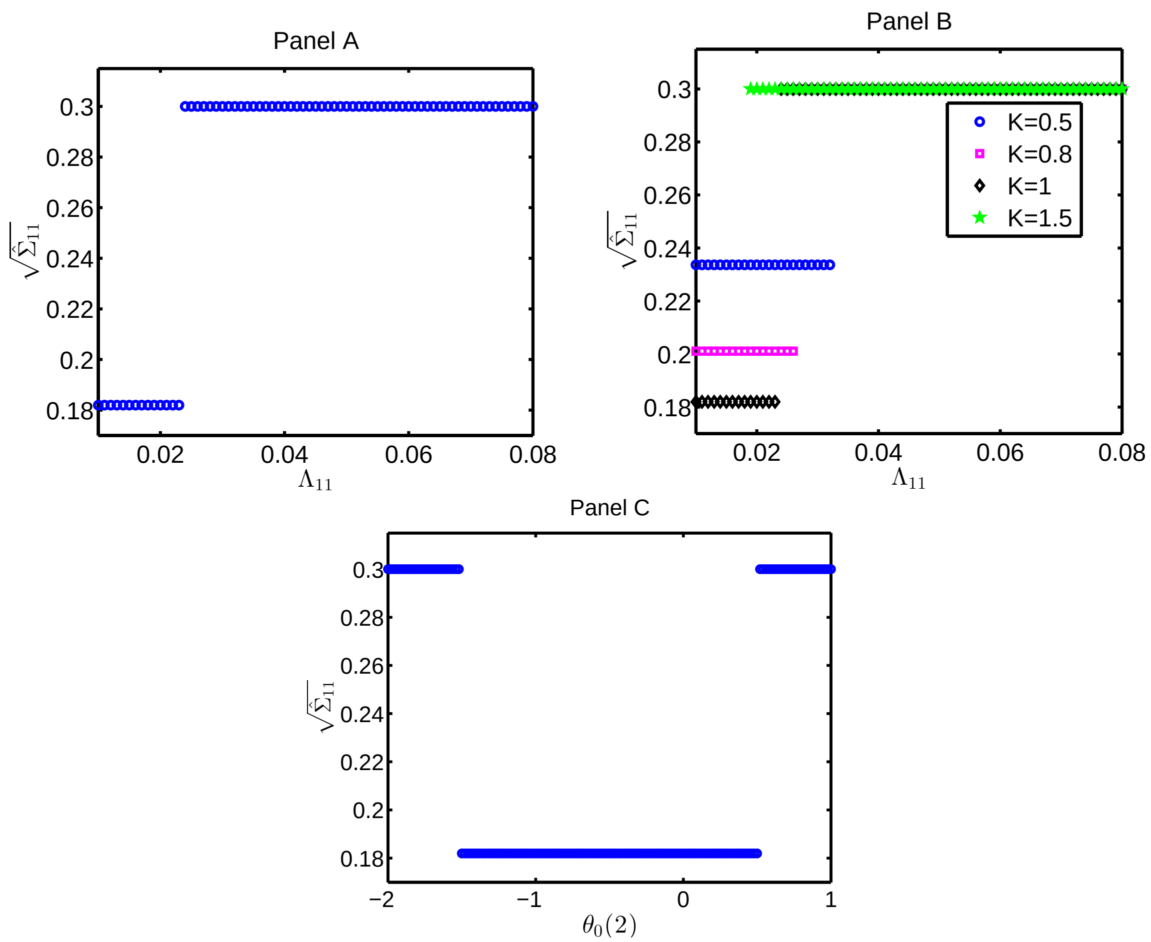

FIGURE 4.5: *Information acquisition with quadratic transaction costs and an entropy learning technology: Specialized learning*

This figure describes investors' optimal information acquisition in a setting in which specialized learning arises. The figure reports the optimal choice of learning about asset 1 measured in terms of posterior volatility $\sqrt{\hat{\mathbf{\Sigma}}_{11}}$. Panel A shows the case in which the transaction costs of asset 1 vary. Panel B extends Panel A by allowing for various learning capacities. Panel C plots the case in which the initial allocation of asset 2 changes. The parameters for Panel A are $\mu_1 - rP_1 = 0.1$, $\mu_2 - rP_2 = 0.05$, $\sqrt{\mathbf{\Sigma}_{11}} = 0.3$, $\sqrt{\mathbf{\Sigma}_{22}} = 0.3$, $\Lambda_{22} = 0.01$, $K = 1$, $\theta^0(1) = 0$, $\theta^0(2) = 0$ and $\gamma = 5$. In Panel B, the parameter values are the same as those in Panel A, and we allow learning capacity to vary from $K = 0.5$ to 0.8, 1 and 1.5. The parameters for Panel C are $\mu_1 - rP_1 = 0.07$, $\mu_2 - rP_2 = 0.05$, $\sqrt{\mathbf{\Sigma}_{11}} = 0.3$, $\sqrt{\mathbf{\Sigma}_{22}} = 0.3$, $\Lambda_{11} = 0.1$, $\Lambda_{22} = 0.1$, $K = 1$, $\theta^0(1) = 0.3$ and $\gamma = 5$.

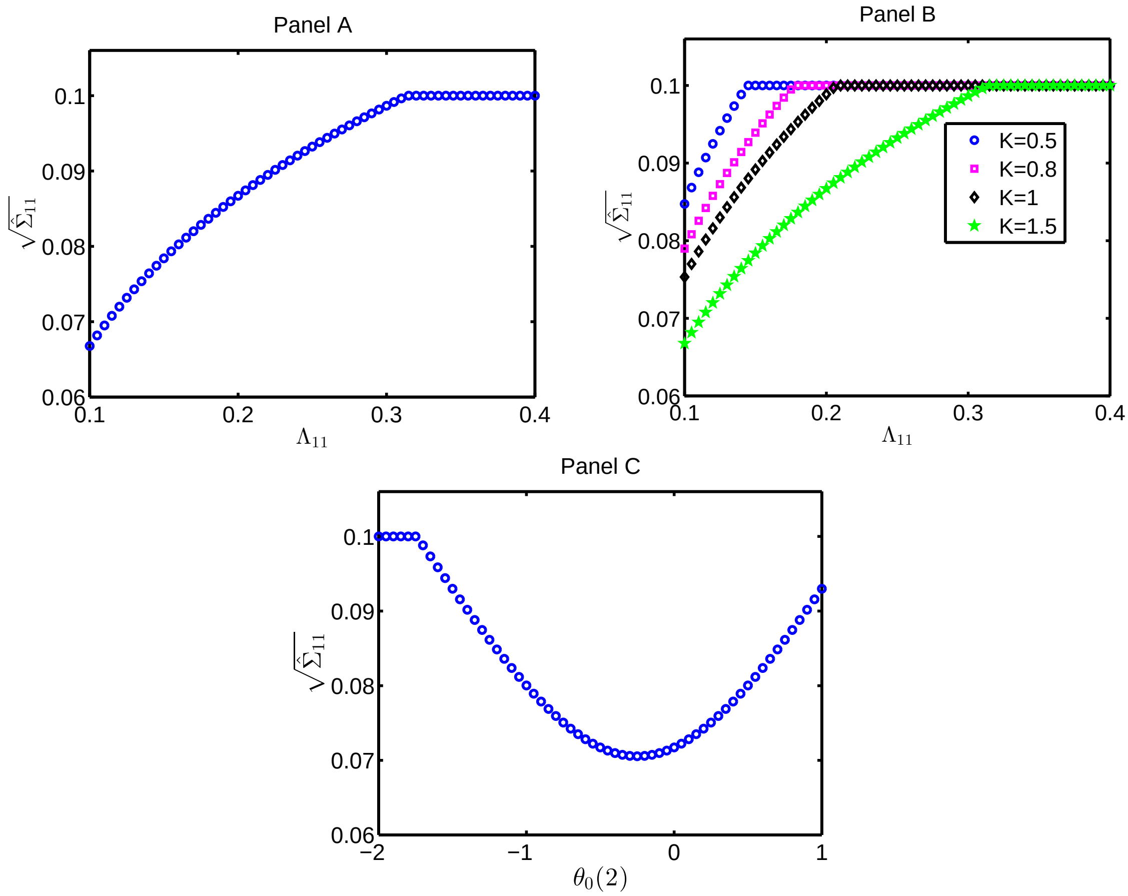

FIGURE 4.6: *Information acquisition with quadratic transaction costs and an entropy learning technology: Generalized learning*

This figure describes investors' optimal information acquisition in a setting in which generalized learning arises. The figure reports the optimal choice of learning about asset 1 measured in terms of posterior volatility $\sqrt{\hat{\mathbf{\Sigma}}_{11}}$. Panel A shows the case in which the transaction costs of asset 1 vary. Panel B extends Panel A by allowing for various learning capacities. Panel C plots the case in which the initial allocation of asset 2 changes. The parameters for Panel A are $\mu_1 - rP_1 = 0.1$, $\mu_2 - rP_2 = 0.05$, $\sqrt{\mathbf{\Sigma}_{11}} = 0.1$, $\sqrt{\mathbf{\Sigma}_{22}} = 0.1$, $\Lambda_{22} = 0.1$, $K = 1$, $\theta^0(1) = 0$, $\theta^0(2) = 0$ and $\gamma = 5$. In Panel B, the parameter values are the same as those in Panel A, and we allow learning capacity to vary from $K = 0.5$ to 0.8, 1 and 1.5. The parameters for Panel C are $\mu_1 - rP_1 = 0.07$, $\mu_2 - rP_2 = 0.05$, $\sqrt{\mathbf{\Sigma}_{11}} = 0.1$, $\sqrt{\mathbf{\Sigma}_{22}} = 0.1$, $\Lambda_{11} = 0.2$, $\Lambda_{22} = 0.2$, $K = 1$, $\theta^0(1) = 0.3$ and $\gamma = 5$.

4.5 Conclusion

Taking transaction costs into account has a large impact on optimal portfolios and asset prices. The literature has recognized that transaction costs are a significant factor in determining investors' trading and consumption behavior. However, the impact of transaction costs on investors' optimal information acquisition is seldom discussed in the literature. This paper is one of the first attempts made in this respect.

We conduct our analysis in a simple setting with mean-variance investors and consider both proportional and quadratic transaction costs. As expected, we find that liquidity (measured by transaction costs) affects the relative attractiveness of learning about asset payoffs to investors. As an asset's transaction costs increase, this asset becomes less attractive to learn about. Moreover, we find that investors with different learning capacities and initial allocations choose to learn about different assets. The most striking finding is that transaction costs might also affect whether investors choose specialized or generalized learning. Hence, once one accounts for liquidity, several factors besides assets' Sharpe ratio play an important role in determining investors' learning behavior.

APPENDIX

4.A Portfolio Choice with Proportional Transaction Costs

This section presents the solution to the portfolio choice problem discussed in Section 4.2.4. Our starting point is the Lagrangian function and the first-order condition in equations (4.12) and (4.15) shown in the text. Although we assume in the text that the two risky assets are independent under the posterior belief, we present the solution for the general case and denote the correlation coefficient by ρ.

The Lagrangian multiplier corresponding to the non-zero elements of the trading vector $\boldsymbol{q}$ must be zero. To determine which elements of $\boldsymbol{q}$ are non-zero, we use the first-order condition (4.15) with $\boldsymbol{\lambda}$ set to $\mathbf{0}$:

$$\boldsymbol{T}'\left(\hat{\boldsymbol{\mu}} - \boldsymbol{P}r - \gamma\hat{\boldsymbol{\Sigma}}\boldsymbol{\theta}^0\right) - \gamma\boldsymbol{T}'\hat{\boldsymbol{\Sigma}}\boldsymbol{T}\boldsymbol{q} - \boldsymbol{C} = \mathbf{0}\,.$$

We let $\boldsymbol{V}$ denote the matrix $\boldsymbol{T}'\hat{\boldsymbol{\Sigma}}\boldsymbol{T}$. Assuming

$$\hat{\Sigma} = \begin{bmatrix} \hat{\sigma}_1^2 & \rho\hat{\sigma}_1\hat{\sigma}_2 \\ \rho\hat{\sigma}_1\hat{\sigma}_2 & \hat{\sigma}_2^2 \end{bmatrix},$$

the elements of $\boldsymbol{V}$, v_{ij} can be written as

$$v_{ij} = T_{1i}T_{1j}\hat{\sigma}_1^2 + (T_{2i}T_{1j} + T_{1i}T_{2j})\rho\hat{\sigma}_1\hat{\sigma}_2 + T_{2i}T_{2j}\hat{\sigma}_2^2\,.$$

Define x_1 and x_2 as follows:

$$\begin{aligned} x_1 &= (\hat{\mu}_1 - rP_1) - \gamma\hat{\sigma}_1(\hat{\sigma}_1\theta_1^0 + \rho\hat{\sigma}_2\theta_2^0)\,, \\ x_2 &= (\hat{\mu}_2 - rP_2) - \gamma\hat{\sigma}_2(\rho\hat{\sigma}_1\theta_1^0 + \hat{\sigma}_2\theta_2^0)\,. \end{aligned}$$

To characterize the solution, we must distinguish cases where a single asset is traded and those where two assets are traded. For the cases in which investors trade only a single asset, i.e., only one q_i takes a positive value, then $\lambda_i = 0$ for that i. We obtain the FOC

$$T_{1i}x_i - C_i - \gamma v_{ii}q_i = 0\,.$$

Solving for q_i yields

$$q_i = \frac{T_{1i}x_i - C_i}{\gamma v_{ii}}\,.$$

For the cases in which two assets are traded, i.e. two elements of $\boldsymbol{q}$, q_i and q_j, are positive, the FOC is

$$\begin{bmatrix} T_{1i} & T_{2i} \\ T_{1j} & T_{2j} \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} - \begin{bmatrix} C_i \\ C_j \end{bmatrix} - \gamma \begin{bmatrix} v_{ii} & v_{ij} \\ v_{ji} & v_{jj} \end{bmatrix} \begin{bmatrix} q_i \\ q_j \end{bmatrix} = \mathbf{0} \,, \tag{4.38}$$

Solving for q_i and q_j, we have

$$\begin{bmatrix} q_i \\ q_j \end{bmatrix} = \frac{1}{\gamma(v_{ii}v_{jj} - v_{ij}^2)} \begin{bmatrix} v_{jj} & -v_{ij} \\ -v_{ij} & v_{ii} \end{bmatrix} \begin{bmatrix} T_{1i}x_1 + T_{2i}x_2 - C_i \\ T_{1j}x_1 + T_{2j}x_2 - C_j \end{bmatrix} . \tag{4.39}$$

We introduce some notation to simplify the expressions:

$$A = b[\hat\sigma_1((\hat\mu_1 - rP_1)\hat\sigma_2\rho - (\hat\mu_2 - rP_2)\hat\sigma_1)] + (1-b)[\hat\sigma_2((\hat\mu_1 - rP_1)\hat\sigma_2 - (\hat\mu_2 - rP_2)\rho\hat\sigma_1)] \,,$$

$$\tilde v^1 = b\hat\sigma_1^2 + (1-b)\rho\hat\sigma_1\hat\sigma_2 \,,$$

$$\tilde v^2 = b\rho\hat\sigma_1\hat\sigma_2 + (1-b)\hat\sigma_2^2 \,,$$

$$\tilde v = b^2\hat\sigma_1^2 + 2b(1-b)\rho\hat\sigma_1\hat\sigma_2 + (1-b)^2\hat\sigma_2^2 \,,$$

$$C_1 = \frac{(\hat\mu_1 - rP_1)\hat\sigma_2^2 - (\hat\mu_2 - rP_2)\rho\hat\sigma_1\hat\sigma_2}{\gamma|\hat\Sigma|} \,,$$

$$C_2 = \frac{(\hat\mu_2 - rP_2)\hat\sigma_1^2 - (\hat\mu_1 - rP_1)\rho\hat\sigma_1\hat\sigma_2}{\gamma|\hat\Sigma|} \,,$$

$$\bar\varrho_1 = \frac{\rho\hat\sigma_2(t_2(1-b) - t_3) + t_2 b\hat\sigma_1}{\gamma\hat\sigma_1\hat\sigma_2^2(1-\rho^2)b} = \frac{t_2\tilde v^1 - t_3\rho\hat\sigma_1\hat\sigma_2}{\gamma|\hat\Sigma|b} \,,$$

$$\bar\varrho_2 = \frac{\rho\hat\sigma_1(t_1 b - t_3) + t_1(1-b)\hat\sigma_2}{\gamma\hat\sigma_1^2\hat\sigma_2(1-\rho^2)(1-b)} = \frac{t_1\tilde v^2 - t_3\rho\hat\sigma_1\hat\sigma_2}{\gamma|\hat\Sigma|(1-b)} \,,$$

$$\varrho_1 = \frac{\hat\sigma_2(t_2(1-b) - t_3) + t_2 b\rho\hat\sigma_1}{\gamma\hat\sigma_1^2\hat\sigma_2(1-\rho^2)b} = \frac{t_2\tilde v^2 - t_3\hat\sigma_2^2}{\gamma|\hat\Sigma|b} \,,$$

$$\varrho_2 = \frac{\hat\sigma_1(t_1 b - t_3) + t_1(1-b)\rho\hat\sigma_2}{\gamma\hat\sigma_1\hat\sigma_2^2(1-\rho^2)(1-b)} = \frac{t_1\tilde v^1 - t_3\hat\sigma_1^2}{\gamma|\hat\Sigma|(1-b)} \,,$$

$$\Delta_1 = \frac{t_2\rho\hat\sigma_1\hat\sigma_2 + t_1\hat\sigma_2^2}{\gamma|\hat\Sigma|} \,,$$

$$\Delta_2 = \frac{t_1\rho\hat\sigma_1\hat\sigma_2 + t_2\hat\sigma_1^2}{\gamma|\hat\Sigma|} \,,$$

$$\tilde\kappa_1 = \frac{t_1\tilde v - t_3\tilde v^1}{\gamma|\hat\Sigma|(1-b)} \,,$$

$$\tilde{\kappa}_2 = \frac{t_2 \tilde{v} - t_3 \tilde{v}^2}{\gamma|\hat{\Sigma}|b} ,$$

$$A = b[\hat{\sigma}_1((\hat{\mu}_1 - rP_1)\hat{\sigma}_2\rho - (\hat{\mu}_2 - rP_2)\hat{\sigma}_1)] + (1-b)[\hat{\sigma}_2((\hat{\mu}_1 - rP_1)\hat{\sigma}_2 - (\hat{\mu}_2 - rP_2)\rho\hat{\sigma}_1)] ,$$

$$\tilde{A} = \frac{A}{\gamma|\hat{\Sigma}|} = (1-b)C_1 - bC_2 = \frac{(\hat{\mu}_1 - rP_1)\tilde{v}_2 - (\hat{\mu}_2 - rP_2)\tilde{v}_1}{\gamma|\hat{\Sigma}|} ,$$

$$B = \gamma(\tilde{v}^1\theta_1^0 + \tilde{v}^2\theta_2^0) ,$$

$$\tilde{B} = (b(\hat{\mu}_1 - rP_1) + (1-b)(\hat{\mu}_2 - rP_2)) .$$

We now describe the solutions; it involves a total of 13 cases, corresponding to 12 trade regions and 1 no-trade region. In each case, we describe the non-zero elements of $\boldsymbol{q}$; all elements not mentioned are zero.

Case (1) [sell asset 1; i.e., $q_1 > 0$**]**

$$q_1 = \frac{-(\hat{\mu}_1 - rP_1) - t_1}{\gamma\hat{\sigma}_1^2} + \theta_1^0 + \frac{\rho\hat{\sigma}_2}{\hat{\sigma}_1}\theta_2^0 .$$

Case (2) [buy asset 1; i.e., $q_2 > 0$**]**

$$q_2 = \frac{(\hat{\mu}_1 - rP_1) - t_1}{\gamma\hat{\sigma}_1^2} - \theta_1^0 - \frac{\rho\hat{\sigma}_2}{\hat{\sigma}_1}\theta_2^0 .$$

Case (3) [sell asset 2; i.e., $q_3 > 0$**]**

$$q_3 = \frac{-(\hat{\mu}_2 - rP_2) - t_2}{\gamma\hat{\sigma}_2^2} + \frac{\rho\hat{\sigma}_1}{\hat{\sigma}_2}\theta_1^0 + \theta_2^0 .$$

Case (4) [buy asset 2; i.e., $q_4 > 0$**]**

$$q_4 = \frac{(\hat{\mu}_2 - rP_2) - t_2}{\gamma\hat{\sigma}_2^2} - \frac{\rho\hat{\sigma}_1}{\hat{\sigma}_2}\theta_1^0 - \theta_2^0 .$$

Case (5) [sell bundle; i.e., $q_5 > 0$**]**

$$q_5 = \frac{-(b(\hat{\mu}_1 - rP_1) + (1-b)(\hat{\mu}_2 - rP_2)) - t_3 + \gamma\hat{\sigma}_1(b\hat{\sigma}_1 + (1-b)\rho\hat{\sigma}_2)\theta_1^0 + \gamma\hat{\sigma}_2(b\rho\hat{\sigma}_1 + (1-b)\hat{\sigma}_2)\theta_2^0}{\gamma[b^2\hat{\sigma}_1^2 + 2b(1-b)\hat{\sigma}_1\hat{\sigma}_2 + (1-b)^2\hat{\sigma}_2^2]} .$$

Case (6) [buy bundle; i.e., $q_6 > 0$**]**

$$q_6 = \frac{(b(\hat{\mu}_1 - rP_1) + (1-b)(\hat{\mu}_2 - rP_2)) - t_3 - \gamma\hat{\sigma}_1(b\hat{\sigma}_1 + (1-b)\rho\hat{\sigma}_2)\theta_1^0 - \gamma\hat{\sigma}_2(b\rho\hat{\sigma}_1 + (1-b)\hat{\sigma}_2)\theta_2^0}{\gamma[b^2\hat{\sigma}_1^2 + 2b(1-b)\hat{\sigma}_1\hat{\sigma}_2 + (1-b)^2\hat{\sigma}_2^2]} .$$

Case (7) [sell asset 1, buy asset 2; i.e., $q_1 > 0$, $q_4 > 0$**]**

$$q_1 = \theta_1^0 + \frac{-\hat{\sigma}_2((\hat{\mu}_1 - rP_1)\hat{\sigma}_2 - (\hat{\mu}_2 - rP_2)\rho\hat{\sigma}_1) - t_2\rho\hat{\sigma}_1\hat{\sigma}_2 - t_1\hat{\sigma}_2^2}{\gamma\hat{\sigma}_1^2\hat{\sigma}_2^2(1-\rho^2)},$$

$$q_4 = -\theta_2^0 + \frac{\hat{\sigma}_1((\hat{\mu}_2 - rP_2)\hat{\sigma}_1(\hat{\mu}_1 - rP_1) - \hat{\sigma}_2\rho) - t_1\rho\hat{\sigma}_1\hat{\sigma}_2 - t_2\hat{\sigma}_1^2}{\gamma\hat{\sigma}_1^2\hat{\sigma}_2^2(1-\rho^2)}.$$

Case (8) [buy asset 1, sell asset 2; i.e., $q_2 > 0$, $q_3 > 0$**]**

$$q_2 = -\theta_1^0 + \frac{\hat{\sigma}_2((\hat{\mu}_1 - rP_1)\hat{\sigma}_2 - (\hat{\mu}_2 - rP_2)\rho\hat{\sigma}_1) - t_2\rho\hat{\sigma}_1\hat{\sigma}_2 - t_1\hat{\sigma}_2^2}{\gamma\hat{\sigma}_1^2\hat{\sigma}_2^2(1-\rho^2)},$$

$$q_3 = \theta_2^0 + \frac{-\hat{\sigma}_1((\hat{\mu}_2 - rP_2)\hat{\sigma}_1 - (\hat{\mu}_1 - rP_1)\rho\hat{\sigma}_2) - t_1\rho\hat{\sigma}_1\hat{\sigma}_2 - t_2\hat{\sigma}_1^2}{\gamma\hat{\sigma}_1^2\hat{\sigma}_2^2(1-\rho^2)}.$$

Case (9) [sell asset 1 and bundle; i.e., $q_1 > 0$, $q_5 > 0$**]**

$$q_1 = -\frac{(b-1)\theta_1^0 + b\theta_2^0}{1-b} + \frac{-(1-b)A + t_3\tilde{v}^1 - t_1\tilde{v}}{\gamma\hat{\sigma}_1^2\hat{\sigma}_2^2(1-\rho^2)(1-b)^2},$$

$$q_5 = \frac{\theta_2^0}{1-b} + \frac{(\hat{\mu}_1 - rP_1)\hat{\sigma}_2\rho - (\hat{\mu}_2 - rP_2)\hat{\sigma}_1}{\gamma\hat{\sigma}_1\hat{\sigma}_2^2(1-\rho^2)(1-b)} + \frac{\hat{\sigma}_1(t_1 b - t_3) + t_1(1-b)\rho\hat{\sigma}_2}{\gamma\hat{\sigma}_1\hat{\sigma}_2^2(1-\rho^2)(1-b)^2}.$$

Case (10) [buy asset 1 and bundle; i.e., $q_2 > 0$, $q_6 > 0$**]**

$$q_2 = \frac{b\theta_2^0 - (1-b)\theta_1^0}{1-b} + \frac{(1-b)A + t_3\tilde{v}^1 - t_1\tilde{v}}{\gamma\hat{\sigma}_1^2\hat{\sigma}_2^2(1-\rho^2)(1-b)^2},$$

$$q_6 = -\frac{\theta_2^0}{1-b} + \frac{(\hat{\mu}_2 - rP_2)\hat{\sigma}_1 - (\hat{\mu}_1 - rP_1)\hat{\sigma}_2\rho}{\gamma\hat{\sigma}_1\hat{\sigma}_2^2(1-\rho^2)(1-b)} + \frac{\hat{\sigma}_1(t_1 b - t_3) + t_1(1-b)\rho\hat{\sigma}_2}{\gamma\hat{\sigma}_1\hat{\sigma}_2^2(1-\rho^2)(1-b)^2}.$$

Case (11) [sell asset 2 and bundle; i.e., $q_3 > 0$, $q_5 > 0$**]**

$$q_3 = \frac{[b\theta_2^0 + (b-1)\theta_1^0]}{b} + \frac{bA + t_3\tilde{v}^2 - t_2\tilde{v}}{\gamma\hat{\sigma}_1^2\hat{\sigma}_2^2(1-\rho^2)b^2},$$

$$q_5 = \frac{\theta_1^0}{b} - \frac{(\hat{\mu}_1 - rP_1)\hat{\sigma}_2 - (\hat{\mu}_2 - rP_2)\rho\hat{\sigma}_1}{\gamma\hat{\sigma}_1^2\hat{\sigma}_2(1-\rho^2)b} + \frac{t_2 b\rho\hat{\sigma}_1 - \hat{\sigma}_2(t_3 - t_2(1-b))}{\gamma\hat{\sigma}_1^2\hat{\sigma}_2(1-\rho^2)b^2}.$$

Case (12) [buy asset 2 and bundle; i.e., $q_4 > 0$, $q_6 > 0$**]**

$$q_4 = \frac{(1-b)\theta_1^0 - b\theta_2^0}{b} + \frac{-bA + t_3\tilde{v}^2 - t_2\tilde{v}}{\gamma\hat{\sigma}_1^2\hat{\sigma}_2^2(1-\rho^2)b^2},$$

$$q_6 = -\frac{\theta_1^0}{b} + \frac{(\hat{\mu}_1 - rP_1)\hat{\sigma}_2 - (\hat{\mu}_2 - rP_2)\rho\hat{\sigma}_1}{\gamma\hat{\sigma}_1^2\hat{\sigma}_2(1-\rho^2)b} + \frac{t_2 b\rho\hat{\sigma}_1 - (t_3 - t_2(1-b))\hat{\sigma}_2}{\gamma\hat{\sigma}_1^2\hat{\sigma}_2^2(1-\rho^2)b^2}.$$

4.B Numerical Integration

This appendix shows how to compute the expectation E_1 in equation (4.20). The expectation can be computed as the sum of the integrals of function U over 13 different regions corresponding to the 12 trading regions and the no-trade region. To be more precise,

$$U_1 = E_1(U) = \sum_{i=1}^{13} \int\int_{R_i} U(\hat{\mu}_1, \hat{\mu}_2) f(\hat{\mu}_1, \hat{\mu}_2) d\hat{\mu}_1 d\hat{\mu}_2,$$

where $f(\hat{\mu}_1, \hat{\mu}_2)$ denotes the joint density of $(\hat{\mu}_1, \hat{\mu}_2)$ and R_i denotes the region i. The precise functional form of U differs in the 13 regions and it is derived by inserting the optimal portfolio choice computed in Appendix 4.A into equation (4.11). The boundaries of each region are obtained by requiring the non-zero elements of the trading vector $\boldsymbol{q}$ to be non-negative. The functional form of U and the boundaries of the 13 regions are the following.

Case (1) [sell asset 1]:
The boundary for this region is: $C_2 \geq \theta_2^0 + \varrho_2, \quad C_2 \leq \theta_2^0 + \Delta_2, \quad (\hat{\mu}_1 - rP_1) \leq -t_1 + \gamma\hat{\sigma}_1^2\theta_1^0 + \gamma\rho\hat{\sigma}_1\hat{\sigma}_2\theta_2^0$.
The region exists if $\Delta_2 > \varrho_2$ or $t_3 > t_1 b - t_2(1-b)$.
The utility function is given by:

$$U = \frac{(\hat{\mu}_1 - rP_1)^2}{2\gamma\hat{\sigma_1}^2} + (\hat{\mu}_1 - rP_1)\frac{t_1 - \gamma\theta_2^0\rho\hat{\sigma}_1\hat{\sigma}_2}{\gamma\hat{\sigma}_1^2} + \theta_2^0(\hat{\mu}_2 - rP_2) + \frac{t_1^2}{2\gamma\hat{\sigma}_1^2} - t_1\frac{\theta_1^0\hat{\sigma}_1 + \theta_2^0\rho\hat{\sigma}_2}{\hat{\sigma_1}} - \frac{\gamma}{2}(\theta_2^0)^2\sigma_2^2(1-\rho^2) .$$

The integration is performed over the following region:
$(\hat{\mu}_1 - rP_1) \in [-\infty, -t_1 + \gamma\hat{\sigma}_1^2\theta_1^0 + \gamma\rho\hat{\sigma}_1\hat{\sigma}_2\theta_2^0]$,
$(\hat{\mu}_2 - rP_2) \in [\frac{(\theta_2^0 + \varrho_2)\gamma|\hat{\Sigma}|}{\hat{\sigma}_1^2} + (\hat{\mu}_1 - rP_1)\frac{\rho\hat{\sigma}_2}{\hat{\sigma}_1}, \frac{(\theta_2^0 + \Delta_2)\gamma|\hat{\Sigma}|}{\hat{\sigma}_1^2} + (\hat{\mu}_1 - rP_1)\frac{\rho\hat{\sigma}_2}{\hat{\sigma}_1}]$.

Case (2) [buy asset 1]:

The boundary for this region is: $C_2 \leq \theta_2^0 - \varrho_2, \quad C_2 \geq \theta_2^0 - \Delta_2, \quad (\hat{\mu}_1 - rP_1) \geq t_1 + \gamma\hat{\sigma}_1^2\theta_1^0 + \gamma\rho\hat{\sigma}_1\hat{\sigma}_2\theta_2^0$.
The region exists if $\Delta_2 > \varrho_2$ or $t_3 > t_1 b - t_2(1-b)$.
The utility function is given by:

$$U = \frac{(\hat{\mu}_1 - rP_1)^2}{2\gamma\hat{\sigma_1}^2} + (\hat{\mu}_1 - rP_1)\frac{-t_1 - \gamma\theta_2^0\rho\hat{\sigma}_1\hat{\sigma}_2}{\gamma\hat{\sigma}_1^2} + \theta_2^0(\hat{\mu}_2 - rP_2) + \frac{t_1^2}{2\gamma\hat{\sigma}_1^2} + t_1\frac{\theta_1^0\hat{\sigma}_1 + \theta_2^0\rho\hat{\sigma}_2}{\hat{\sigma_1}} - \frac{\gamma}{2}(\theta_2^0)^2\sigma_2^2(1-\rho^2) .$$

The integration is performed over the following region:
$(\hat{\mu}_1 - rP_1) \in [t_1 + \gamma\hat{\sigma}_1^2\theta_1^0 + \gamma\rho\hat{\sigma}_1\hat{\sigma}_2\theta_2^0, +\infty]$,
$(\hat{\mu}_2 - rP_2) \in [\frac{(\theta_2^0 - \Delta_2)\gamma|\hat{\Sigma}|}{\hat{\sigma_1}^2} + (\hat{\mu}_1 - rP_1)\frac{\rho\hat{\sigma}_2}{\hat{\sigma}_1}, \frac{(\theta_2^0 - \varrho_2)\gamma|\hat{\Sigma}|}{\hat{\sigma}_1^2} + (\hat{\mu}_1 - rP_1)\frac{\rho\hat{\sigma}_2}{\hat{\sigma}_1}]$.

Case (3) [sell asset 2]:

The boundary for this region is: $C_1 \geq \theta_1^0 + \varrho_1, \quad C_1 \leq \theta_1^0 + \Delta_1, \quad (\hat{\mu}_2 - rP_2) \leq -t_2 + \gamma\hat{\sigma}_2^2\theta_2^0 + \gamma\rho\hat{\sigma}_1\hat{\sigma}_2\theta_1^0$.
The region exists if $\Delta_1 > \varrho_1$ or $t_3 > t_2(1-b) - t_1 b$.
The utility function is given by:

$$U = \frac{(\hat{\mu}_2 - rP_2)^2}{2\gamma\hat{\sigma_2}^2} + (\hat{\mu}_2 - rP_2)\frac{t_2 - \gamma\theta_1^0\rho\hat{\sigma}_1\hat{\sigma}_2}{\gamma\hat{\sigma}_2^2} + \theta_1^0(\hat{\mu}_1 - rP_1) + \frac{t_2^2}{2\gamma\hat{\sigma}_2^2} - t_2\frac{\theta_2^0\hat{\sigma}_2 + \theta_1^0\rho\hat{\sigma}_1}{\hat{\sigma_2}} - \frac{\gamma}{2}(\theta_1^0)^2\sigma_1^2(1-\rho^2) .$$

The integration is performed over the following region:
$(\hat{\mu}_2 - rP_2) \in [-\infty, -t_2 + \gamma\hat{\sigma}_2^2\theta_2^0 + \gamma\rho\hat{\sigma}_1\hat{\sigma}_2\theta_1^0]$.
$(\hat{\mu}_1 - rP_1) \in [\frac{(\theta_1^0 + \varrho_1)\gamma|\hat{\Sigma}|}{\sigma_2^2} + (\hat{\mu}_2 - rP_2)\frac{\rho\hat{\sigma}_1}{\hat{\sigma}_2}, \frac{(\theta_1^0 + \Delta_1)\gamma|\hat{\Sigma}|}{\sigma_2^2} + (\hat{\mu}_2 - rP_2)\frac{\rho\hat{\sigma}_1}{\hat{\sigma}_2}]$.

Case (4) [buy asset 2]:

The boundary for this region is: $C_1 \le \theta_1^0 - \varrho_1, \quad C_1 \ge \theta_1^0 - \Delta_1, \quad (\hat{\mu}_2 - rP_2) \ge t_2 + \gamma\hat{\sigma}_2^2\theta_2^0 + \gamma\rho\hat{\sigma}_1\hat{\sigma}_2\theta_1^0$.
The region exists if $\Delta_1 > \varrho_1$ or $t_3 > t_2(1-b) - t_1 b$.
The utility function is given by:

$$U = \frac{(\hat{\mu}_2 - rP_2)^2}{2\gamma\hat{\sigma_2}^2} + (\hat{\mu}_2 - rP_2)\frac{-t_2 - \gamma\theta_1^0\rho\hat{\sigma}_1\hat{\sigma}_2}{\gamma\hat{\sigma}_2^2} + \theta_1^0(\hat{\mu}_1 - rP_1) + \frac{t_2^2}{2\gamma\hat{\sigma}_2^2} + t_2\frac{\theta_2^0\hat{\sigma}_2 + \theta_1^0\rho\hat{\sigma}_1}{\hat{\sigma_2}} - \frac{\gamma}{2}(\theta_1^0)^2\sigma_1^2(1-\rho^2) .$$

The integration is performed over the following region:
$(\hat{\mu}_2 - rP_2) \in [t_2 + \gamma\hat{\sigma}_2^2\theta_2^0 + \gamma\rho\hat{\sigma}_1\hat{\sigma}_2\theta_1^0, +\infty]$.
$(\hat{\mu}_1 - rP_1) \in [\frac{(\theta_1^0 - \Delta_1)\gamma|\hat{\Sigma}|}{\hat{\sigma}_2^2} + (\hat{\mu}_2 - rP_2)\frac{\rho\hat{\sigma}_1}{\hat{\sigma}_2}, \frac{(\theta_1^0 - \varrho_1)\gamma|\hat{\Sigma}|}{\hat{\sigma}_2^2} + (\hat{\mu}_2 - rP_2)\frac{\rho\hat{\sigma}_1}{\hat{\sigma}_2}]$.

Case (5) [sell bundle]:

The boundary for this region is: $\tilde{A} \ge \theta_1^0(1-b) - \theta_2^0 b - \tilde{\kappa}_1, \quad \tilde{A} \le \theta_1^0(1-b) - \theta_2^0 b + \tilde{\kappa}_2, \quad \tilde{B} \le B - t_3$.
The region exists if $\tilde{\kappa}_2 > -\tilde{\kappa}_1$ or $bt_1 + (1-b)t_2 > t_3$.
The utility function is given by:

$$U = \frac{\tilde{B}^2}{2\gamma\tilde{v}} + \frac{(\hat{\mu}_1 - rP_1)(\gamma((1-b)\theta_1^0 - b\theta_2^0)\tilde{v}_2 + t_3 b) - (\hat{\mu}_2 - rP_2)(\gamma((1-b)\theta_1^0 - b\theta_2^0)\tilde{v}_1 - t_3(1-b))}{\gamma\tilde{v}}$$
$$+ \left[\frac{t_3^2 - 2\gamma t_3(\tilde{v}_1\theta_1^0 + \tilde{v}_2\theta_2^0)}{2\gamma\tilde{v}} - \frac{\gamma((1-b)\theta_1^0 - b\theta_2^0)^2|\hat{\Sigma}|}{2\tilde{v}}\right] .$$

Define

$$(\hat{\mu}_1 - rP_1)^* = \frac{(\gamma(\tilde{v}_1\theta_1^0 + \tilde{v}_2\theta_2^0) - t_3)\tilde{v}_1 + ((\theta_1^0(1-b) - \theta_2^0 b)\gamma|\hat{\Sigma}|)(1-b) - t_1\tilde{v} + t_3\tilde{v}_1}{\tilde{v}} ,$$

$$(\hat{\mu}_1 - rP_1)^\dagger = (\hat{\mu}_1 - rP_1)^* + \frac{t_1 b + t_2(1-b) - t_3}{b} .$$

The integral is the sum of integrals over two sub-regions. The first integral is computed over the following region:
$(\hat{\mu}_1 - rP_1) \in [-\infty, (\hat{\mu}_1 - rP_1)^*]$,
$(\hat{\mu}_2 - rP_2) \in \left[\frac{(\hat{\mu}_1 - rP_1)\tilde{v}_2 - (\theta_1^0(1-b) - \theta_2^0 b)\gamma|\hat{\Sigma}|}{\tilde{v}_1} - \frac{t_2\tilde{v} - t_3\tilde{v}_2}{b\tilde{v}_1}, \frac{(\hat{\mu}_1 - rP_1)\tilde{v}_2 - (\theta_1^0(1-b) - \theta_2^0 b)\gamma|\hat{\Sigma}|}{\tilde{v}_1} + \frac{t_1\tilde{v} - t_3\tilde{v}_1}{(1-b)\tilde{v}_1}\right]$.
The second integral is computed over the following region:
$(\hat{\mu}_1 - rP_1) \in [(\hat{\mu}_1 - rP_1)^*, (\hat{\mu}_1 - rP_1)^\dagger]$,
$(\hat{\mu}_2 - rP_2) \in \left[\frac{(\hat{\mu}_1 - rP_1)\tilde{v}_2 - (\theta_1^0(1-b) - \theta_2^0 b)\gamma|\hat{\Sigma}|}{\tilde{v}_1} - \frac{t_2\tilde{v} - t_3\tilde{v}_2}{b\tilde{v}_1}, \frac{\gamma(\tilde{v}_1\theta_1^0 + \tilde{v}_2\theta_2^0) - t_3 - b(\hat{\mu}_1 - rP_1)}{1-b}\right]$.

Case (6) [buy bundle]:

The boundary for this region is: $\tilde{A} \le \theta_1^0(1-b) - \theta_2^0 b + \tilde{\kappa}_1, \quad \tilde{A} \ge \theta_1^0(1-b) - \theta_2^0 b - \tilde{\kappa}_2, \quad \tilde{B} \ge B + t_3$.
The region exists if $\tilde{\kappa}_2 > -\tilde{\kappa}_1$ or $bt_1 + (1-b)t_2 > t_3$.
The utility function is given by:

$$
\begin{aligned}
U &= \frac{\tilde{B}^2}{2\gamma\tilde{v}} + \frac{(\hat{\mu}_1 - rP_1)(\gamma((1-b)\theta_1^0 - b\theta_2^0)\tilde{v}_2 - t_3 b) - (\hat{\mu}_2 - rP_2)(\gamma((1-b)\theta_1^0 - b\theta_2^0)\tilde{v}_1 + t_3(1-b))}{\gamma\tilde{v}} \\
&\quad + \left[\frac{t_3^2 + 2\gamma t_3(\tilde{v}_1\theta_1^0 + \tilde{v}_2\theta_2^0)}{2\gamma\tilde{v}} - \frac{\gamma((1-b)\theta_1^0 - b\theta_2^0)^2|\hat{\Sigma}|}{2\tilde{v}}\right].
\end{aligned}
$$

Denote

$$
(\hat{\mu}_1 - rP_1)^* = \frac{(\gamma(\tilde{v}_1\theta_1^0 + \tilde{v}_2\theta_2^0) + t_3)\tilde{v}_1 + ((\theta_1^0(1-b) - \theta_2^0 b)\gamma|\hat{\Sigma}|)(1-b) + t_1\tilde{v} - t_3\tilde{v}_1}{\tilde{v}},
$$

$$
(\hat{\mu}_1 - rP_1)^\dagger = (\hat{\mu}_1 - rP_1)^* - \frac{t_1 b + t_2(1-b) - t_3}{b}.
$$

The integral is the sum of integrals over two sub-regions. The first integral is computed over the following region:

$(\hat{\mu}_1 - rP_1) \in [(\hat{\mu}_1 - rP_1)^\dagger, (\hat{\mu}_1 - rP_1)^*]$,

$(\hat{\mu}_2 - rP_2) \in \left[\frac{\gamma(\tilde{v}_1\theta_1^0 + \tilde{v}_2\theta_2^0) + t_3 - b(\hat{\mu}_1 - rP_1)}{1-b}, \frac{(\hat{\mu}_1 - rP_1)\tilde{v}_2 - (\theta_1^0(1-b) - \theta_2^0 b)\gamma|\hat{\Sigma}|}{\tilde{v}_1} + \frac{t_2\tilde{v} - t_3\tilde{v}_2}{b\tilde{v}_1}\right]$.

The second integral is computed over the following region:

$(\hat{\mu}_1 - rP_1) \in [(\hat{\mu}_1 - rP_1)^*, +\infty]$,

$(\hat{\mu}_2 - rP_2) \in \left[\frac{(\hat{\mu}_1 - rP_1)\tilde{v}_2 - (\theta_1^0(1-b) - \theta_2^0 b)\gamma|\hat{\Sigma}|}{\tilde{v}_1} - \frac{t_1\tilde{v} - t_3\tilde{v}_1}{(1-b)\tilde{v}_1}, \frac{(\hat{\mu}_1 - rP_1)\tilde{v}_2 - (\theta_1^0(1-b) - \theta_2^0 b)\gamma|\hat{\Sigma}|}{\tilde{v}_1} + \frac{t_2\tilde{v} - t_3\tilde{v}_2}{b\tilde{v}_1}\right]$.

Case (7) [sell asset 1, buy asset 2]:

The boundary for this region is: $C_1 \le \theta_1^0 - \Delta_1, \quad C_2 \ge \theta_2^0 + \Delta_2$.

The utility function is given by:

$$
U = \frac{(\hat{\mu}_1 - rP_1)^2\hat{\sigma}_2^2 - 2(\hat{\mu}_1 - rP_1)(\hat{\mu}_2 - rP_2)\rho\hat{\sigma}_1\hat{\sigma}_2 + (\hat{\mu}_2 - rP_2)^2\hat{\sigma}_1^2}{2\gamma|\hat{\Sigma}|} - (\hat{\mu}_2 - rP_2)\Delta_2 + (\hat{\mu}_1 - rP_1)\Delta_1
$$

$$
+ \frac{t_1^2\hat{\sigma}_2^2 + 2t_1t_2\rho\hat{\sigma}_1\hat{\sigma}_2 + t_2^2\hat{\sigma}_1^2}{2\gamma|\hat{\Sigma}|} - t_1\theta_1^0 + t_2\theta_2^0.
$$

Define:

$$
(\hat{\mu}_1 - rP_1)^* = (\rho\hat{\sigma}_1\hat{\sigma}_2\theta_2^0 + \hat{\sigma}_1^2\theta_1^0)\gamma - t_1.
$$

The integral is the sum of integrals over two sub-regions. The first integral is computed over the following region:

$(\hat{\mu}_1 - rP_1) \in [-\infty, (\hat{\mu}_1 - rP_1)^*]$,

$(\hat{\mu}_2 - rP_2) \in [(\hat{\mu}_1 - rP_1)\frac{\rho\hat{\sigma}_2}{\hat{\sigma}_1} + \frac{(\theta_2^0 + \Delta_2)\gamma|\hat{\Sigma}|}{\hat{\sigma}_1^2}, +\infty]$.

The second integral is computed over the following region:

$(\hat{\mu}_1 - rP_1) \in [(\hat{\mu}_1 - rP_1)^*, +\infty]$,

$(\hat{\mu}_2 - rP_2) \in [(\hat{\mu}_1 - rP_1)\frac{\hat{\sigma}_2}{\rho\hat{\sigma}_1} - \frac{(\theta_1^0 - \Delta_1)\gamma|\hat{\Sigma}|}{\rho\hat{\sigma}_1\hat{\sigma}_2}, +\infty]$.

Case (8) [buy asset 1, sell asset 2]:

The boundary for this region is: $C_1 \ge \theta_1^0 + \Delta_1, \quad C_2 \le \theta_2^0 - \Delta_2$.

The utility function is given by:

$$U = \frac{(\hat{\mu}_1 - rP_1)^2\hat{\sigma}_2^2 - 2(\hat{\mu}_1 - rP_1)(\hat{\mu}_2 - rP_2)\rho\hat{\sigma}_1\hat{\sigma}_2 + (\hat{\mu}_2 - rP_2)^2\hat{\sigma}_1^2}{2\gamma|\hat{\Sigma}|} + (\hat{\mu}_2 - rP_2)\Delta_2 - (\hat{\mu}_1 - rP_1)\Delta_1$$

$$+ \frac{t_1^2\hat{\sigma}_2^2 + 2t_1t_2\rho\hat{\sigma}_1\hat{\sigma}_2 + t_2^2\hat{\sigma}_1^2}{2\gamma|\hat{\Sigma}|} + t_1\theta_1^0 - t_2\theta_2^0 .$$

Denote

$$(\hat{\mu}_1 - rP_1)^* = (\rho\hat{\sigma}_1\hat{\sigma}_2\theta_2^0 + \hat{\sigma}_1^2\theta_1^0)\gamma + t_1 .$$

The integral is the sum of integrals over two sub-regions. The first integral is computed over the following region:
$(\hat{\mu}_1 - rP_1) \in [-\infty, (\hat{\mu}_1 - rP_1)^*]$,
$(\hat{\mu}_2 - rP_2) \in [-\infty, (\hat{\mu}_1 - rP_1)\dfrac{\hat{\sigma}_2}{\rho\hat{\sigma}_1} - \dfrac{(\theta_1^0 + \Delta_1)\gamma|\hat{\Sigma}|}{\rho\hat{\sigma}_1\hat{\sigma}_2}]$.
The second integral is computed over the following region:
$(\hat{\mu}_1 - rP_1) \in [(\hat{\mu}_1 - rP_1)^*, +\infty]$,
$(\hat{\mu}_2 - rP_2) \in [-\infty, (\hat{\mu}_1 - rP_1)\dfrac{\rho\hat{\sigma}_2}{\hat{\sigma}_1} + \dfrac{(\theta_2^0 - \Delta_2)\gamma|\hat{\Sigma}|}{\hat{\sigma}_1^2}]$.
Case (9) [sell asset 1, sell bundle]:
The boundary for this region is: $C_2 \le \theta_2^0 + \varrho_2, \quad \tilde{A} \le \theta_1^0(1-b) - \theta_2^0 b - \tilde{\kappa}_1$.
The utility function is given by:

$$U = \frac{(\hat{\mu}_1 - rP_1)^2\hat{\sigma}_2^2 - 2(\hat{\mu}_1 - rP_1)(\hat{\mu}_2 - rP_2)\rho\hat{\sigma}_1\hat{\sigma}_2 + (\hat{\mu}_2 - rP_2)^2\hat{\sigma}_1^2}{2\gamma|\hat{\Sigma}|} + (\hat{\mu}_1 - rP_1)\bar{\varrho}_2 - (\hat{\mu}_2 - rP_2)\varrho_2$$

$$+ \frac{t_3^2\hat{\sigma}_1^2 - 2t_3t_1\tilde{v}_1 + t_1^2\tilde{v}}{2(1-b)^2\gamma|\hat{\Sigma}|} - t_1\theta_1^0 - \frac{\theta_2^0(t_3 - bt_1)}{1-b} .$$

The integration is performed over the following region:
$(\hat{\mu}_1 - rP_1) \in [-\infty, \dfrac{(\theta_2^0 + \varrho_2)\gamma\tilde{v}_1}{1-b} + \dfrac{(\theta_1^0(1-b) - \theta_2^0 b - \tilde{\kappa}_1)\gamma\sigma_1^2}{1-b}]$,
$(\hat{\mu}_2 - rP_2) \in \left[\dfrac{(\hat{\mu}_1 - rP_1)\tilde{v}_2}{\tilde{v}_1} - \dfrac{(\theta_1^0(1-b) - \theta_2^0 b - \tilde{\kappa}_1)\gamma|\hat{\Sigma}|}{\tilde{v}_1}, \dfrac{(\hat{\mu}_1 - rP_1)\rho\hat{\sigma}_2}{\hat{\sigma}_1} + \dfrac{(\theta_2^0 + \varrho_2)\gamma|\hat{\Sigma}|}{\hat{\sigma}_1^2}\right]$.
Case (10) [buy asset 1, buy bundle]:

The boundary for this region is: $C_2 \ge \theta_2^0 - \varrho_2, \quad \tilde{A} \ge \theta_1^0(1-b) - \theta_2^0 b + \tilde{\kappa}_1$.
The utility function is given by:

$$U = \frac{(\hat{\mu}_1 - rP_1)^2\hat{\sigma}_2^2 - 2(\hat{\mu}_1 - rP_1)(\hat{\mu}_2 - rP_2)\rho\hat{\sigma}_1\hat{\sigma}_2 + (\hat{\mu}_2 - rP_2)^2\hat{\sigma}_1^2}{2\gamma|\hat{\Sigma}|} - (\hat{\mu}_1 - rP_1)\bar{\varrho}_2 + (\hat{\mu}_2 - rP_2)\varrho_2$$

$$+ \frac{t_3^2\hat{\sigma}_1^2 - 2t_3t_1\tilde{v}_1 + t_1^2\tilde{v}}{2(1-b)^2\gamma|\hat{\Sigma}|} + t_1\theta_1^0 + \frac{\theta_2^0(t_3 - bt_1)}{1-b} .$$

The integration is performed over the following region:
$(\hat{\mu}_1 - rP_1) \in [\dfrac{(\theta_2^0 - \varrho_2)\gamma\tilde{v}_1}{1-b} + \dfrac{(\theta_1^0(1-b) - \theta_2^0 b + \tilde{\kappa}_1)\gamma\sigma_1^2}{1-b}, +\infty]$.
$(\hat{\mu}_2 - rP_2) \in \left[\dfrac{(\hat{\mu}_1 - rP_1)\rho\hat{\sigma}_2}{\hat{\sigma}_1} + \dfrac{(\theta_2^0 - \varrho_2)\gamma|\hat{\Sigma}|}{\hat{\sigma}_1^2}, \dfrac{(\hat{\mu}_1 - rP_1)\tilde{v}_2}{\tilde{v}_1} - \dfrac{(\theta_1^0(1-b) - \theta_2^0 b + \tilde{\kappa}_1)\gamma|\hat{\Sigma}|}{\tilde{v}_1}\right]$.

Case (11) [sell asset 2, sell bundle]:

The boundary for this region is: $C_1 \le \theta_1^0 + \varrho_1, \quad \tilde{A} \ge \theta_1^0(1-b) - \theta_2^0 b + \tilde{\kappa}_2$.
The utility function is given by:

$$U = \frac{(\hat{\mu}_1 - rP_1)^2\hat{\sigma}_2^2 - 2(\hat{\mu}_1 - rP_1)(\hat{\mu}_2 - rP_2)\rho\hat{\sigma}_1\hat{\sigma}_2 + (\hat{\mu}_2 - rP_2)^2\hat{\sigma}_1^2}{2\gamma|\hat{\Sigma}|} + (\hat{\mu}_2 - rP_2)\bar{\varrho}_1 - (\hat{\mu}_1 - rP_1)\varrho_1$$

$$+ \frac{t_3^2\hat{\sigma}_2^2 - 2t_3t_2\tilde{v}_2 + t_2^2\tilde{v}}{2b^2\gamma|\hat{\Sigma}|} - t_2\theta_2^0 - \frac{\theta_1^0(t_3 - (1-b)t_2)}{b} .$$

The integration is performed over the following region:
$(\hat{\mu}_2 - rP_2) \in [-\infty, \frac{(\theta_1^0 + \varrho_1)\gamma\tilde{v}_2}{b} + \frac{(\theta_2^0 b - \theta_1^0(1-b) - \tilde{\kappa}_2)\gamma\sigma_2^2}{b}]$,
$(\hat{\mu}_1 - rP_1) \in \left[\frac{(\hat{\mu}_2 - rP_2)\tilde{v}_1}{\tilde{v}_2} + \frac{(-\theta_2^0 b + \theta_1^0(1-b) + \tilde{\kappa}_2)\gamma|\hat{\Sigma}|}{\tilde{v}_2}, \frac{\rho(\hat{\mu}_2 - rP_2)\hat{\sigma}_1}{\hat{\sigma}_2} + \frac{(\theta_1^0 + \varrho_1)\gamma|\hat{\Sigma}|}{\hat{\sigma}_2^2}\right]$.

Case (12) [buy asset 2, buy bundle]:

The boundary for this region is: $C_1 \ge \theta_1^0 - \varrho_1, \quad \tilde{A} \le \theta_1^0(1-b) - \theta_2^0 b - \tilde{\kappa}_2$.
The utility function is given by:

$$U = \frac{(\hat{\mu}_1 - rP_1)^2\hat{\sigma}_2^2 - 2(\hat{\mu}_1 - rP_1)(\hat{\mu}_2 - rP_2)\rho\hat{\sigma}_1\hat{\sigma}_2 + (\hat{\mu}_2 - rP_2)^2\hat{\sigma}_1^2}{2\gamma|\hat{\Sigma}|} - (\hat{\mu}_2 - rP_2)\bar{\varrho}_1 + (\hat{\mu}_1 - rP_1)\varrho_1$$

$$+ \frac{t_3^2\hat{\sigma}_2^2 - 2t_3t_2\tilde{v}_2 + t_2^2\tilde{v}}{2b^2\gamma|\hat{\Sigma}|} + t_2\theta_2^0 + \frac{\theta_1^0(t_3 - (1-b)t_2)}{b} .$$

The integration is performed over the following region:
$(\hat{\mu}_2 - rP_2) \in [\frac{(\theta_1^0 - \varrho_1)\gamma\tilde{v}_2}{b} + \frac{(\theta_2^0 b - \theta_1^0(1-b) + \tilde{\kappa}_2)\gamma\sigma_2^2}{b}, +\infty]$,
$(\hat{\mu}_1 - rP_1) \in \left[\frac{(\hat{\mu}_2 - rP_2)\rho\hat{\sigma}_1}{\hat{\sigma}_2} + \frac{(\theta_1^0 - \varrho_1)\gamma|\hat{\Sigma}|}{\hat{\sigma}_2^2}, \frac{(\hat{\mu}_2 - rP_2)\tilde{v}_1}{\tilde{v}_2} - \frac{(\theta_2^0 b - \theta_1^0(1-b) + \tilde{\kappa}_2)\gamma|\hat{\Sigma}|}{\tilde{v}_2}\right]$.

Case (13) [no-trade region]:

The boundary for this region is: $(\hat{\mu}_2 - rP_2) \le t_2 + \gamma\hat{\sigma}_2^2\theta_2^0 + \gamma\rho\hat{\sigma}_1\hat{\sigma}_2\theta_1^0, \quad \tilde{B} \le B + t_3, \quad (\hat{\mu}_1 - rP_1) \le t_1 + \gamma\hat{\sigma}_1^2\theta_1^0 + \gamma\rho\hat{\sigma}_1\hat{\sigma}_2\theta_2^0$,
$(\hat{\mu}_2 - rP_2) \ge -t_2 + \gamma\hat{\sigma}_2^2\theta_2^0 + \gamma\rho\hat{\sigma}_1\hat{\sigma}_2\theta_1^0, \quad \tilde{B} \ge B - t_3, \quad (\hat{\mu}_1 - rP_1) \ge -t_1 + \gamma\hat{\sigma}_1^2\theta_1^0 + \gamma\rho\hat{\sigma}_1\hat{\sigma}_2\theta_2^0$.
The utility function is given by:

$$U = \theta_1^0(\hat{\mu}_1 - rP_1) + \theta_2^0(\hat{\mu}_2 - rP_2) - \frac{\gamma}{2}((\theta_1^0)^2\hat{\sigma}_1^2 + 2\rho\hat{\sigma}_1\hat{\sigma}_2\theta_1^0\theta_2^0 + (\theta_2^0)^2\hat{\sigma}_2^2) .$$

Denote

$$(\hat{\mu}_1 - rP_1)^+ = \frac{t_3}{b} - t_2\frac{1-b}{b} + \gamma(\theta_1^0\hat{\sigma}_1^2 + \theta_2^0\rho\hat{\sigma}_1\hat{\sigma}_2) ,$$

$$(\hat{\mu}_1 - rP_1)^- = t_2\frac{1-b}{b} - \frac{t_3}{b} + \gamma(\theta_1^0\hat{\sigma}_1^2 + \theta_2^0\rho\hat{\sigma}_1\hat{\sigma}_2) .$$

Depending on the level of transaction costs, the integration region differs.
Case 13a: If $t_3 > t_2(1-b)$, we have $(\hat{\mu}_1 - rP_1)^+ > (\hat{\mu}_1 - rP_1)^-$.
The integral is the sum of integrals over three sub-regions. The first integral is computed over the following region:

$(\hat{\mu}_1 - rP_1) \in [-t_1 + \gamma\hat{\sigma}_1^2\theta_1^0 + \gamma\rho\hat{\sigma}_1\hat{\sigma}_2\theta_2^0, (\hat{\mu}_1 - rP_1)^-]$,
$(\hat{\mu}_2 - rP_2) \in [\frac{B - t_3 - b(\hat{\mu}_1 - rP_1)}{1-b}, t_2 + \gamma\hat{\sigma}_2^2\theta_2^0 + \gamma\rho\hat{\sigma}_1\hat{\sigma}_2\theta_1^0]$.
The second integral is computed over the region:
$(\hat{\mu}_1 - rP_1) \in [(\hat{\mu}_1 - rP_1)^-, (\hat{\mu}_1 - rP_1)^+]$,
$(\hat{\mu}_2 - rP_2) \in [-t_2 + \gamma\hat{\sigma}_2^2\theta_2^0 + \gamma\rho\hat{\sigma}_1\hat{\sigma}_2\theta_1^0, t_2 + \gamma\hat{\sigma}_2^2\theta_2^0 + \gamma\rho\hat{\sigma}_1\hat{\sigma}_2\theta_1^0]$.
The third integral is computed over the region:
$(\hat{\mu}_1 - rP_1) \in [(\hat{\mu}_1 - rP_1)^+, t_1 + \gamma\hat{\sigma}_1^2\theta_1^0 + \gamma\rho\hat{\sigma}_1\hat{\sigma}_2\theta_2^0]$,
$(\hat{\mu}_2 - rP_2) \in [-t_2 + \gamma\hat{\sigma}_2^2\theta_2^0 + \gamma\rho\hat{\sigma}_1\hat{\sigma}_2\theta_1^0, \frac{B + t_3 - b(\hat{\mu}_1 - rP_1)}{1-b}]$.
Case 13b: If $t_3 < t_2(1-b)$ holds, we have $(\hat{\mu}_1 - rP_1)^+ < (\hat{\mu}_1 - rP_1)^-$.
The integral is the sum of integrals over three sub-regions. The first integral is computed over the following region:
$(\hat{\mu}_1 - rP_1) \in [-t_1 + \gamma\hat{\sigma}_1^2\theta_1^0 + \gamma\rho\hat{\sigma}_1\hat{\sigma}_2\theta_2^0, (\hat{\mu}_1 - rP_1)^+]$,
$(\hat{\mu}_2 - rP_2) \in [\frac{B - t_3 - b(\hat{\mu}_1 - rP_1)}{1-b}, t_2 + \gamma\hat{\sigma}_2^2\theta_2^0 + \gamma\rho\hat{\sigma}_1\hat{\sigma}_2\theta_1^0]$.
The second integral is computed over the region:
$(\hat{\mu}_1 - rP_1) \in [(\hat{\mu}_1 - rP_1)^+, (\hat{\mu}_1 - rP_1)^-]$,
$(\hat{\mu}_2 - rP_2) \in [\frac{B - t_3 - b(\hat{\mu}_1 - rP_1)}{1-b}, \frac{B + t_3 - b(\hat{\mu}_1 - rP_1)}{1-b}]$.
The third integral is computed over the region:
$(\hat{\mu}_1 - rP_1) \in [(\hat{\mu}_1 - rP_1)^-, t_1 + \gamma\hat{\sigma}_1^2\theta_1^0 + \gamma\rho\hat{\sigma}_1\hat{\sigma}_2\theta_2^0]$,
$(\hat{\mu}_2 - rP_2) \in [-t_2 + \gamma\hat{\sigma}_2^2\theta_2^0 + \gamma\rho\hat{\sigma}_1\hat{\sigma}_2\theta_1^0, \frac{B + t_3 - b(\hat{\mu}_1 - rP_1)}{1-b}]$.

4.C Multiple Assets and Quadratic Transaction Costs

Substituting the optimal portfolio choice in equation (4.27) back into the objective function in the second period as defined in equation (4.25) yields

$$\begin{aligned} U &= \left(\hat{\boldsymbol{\mu}} - r\boldsymbol{P} + \boldsymbol{\Lambda}\boldsymbol{\theta}^{\mathbf{0}}\right)' \boldsymbol{A}^{-1}(\hat{\boldsymbol{\mu}} - r\boldsymbol{P}) - \frac{\gamma}{2}\left(\hat{\boldsymbol{\mu}} - r\boldsymbol{P} + \boldsymbol{\Lambda}\boldsymbol{\theta}^{\mathbf{0}}\right)' \boldsymbol{A}^{-1}\hat{\boldsymbol{\Sigma}}\boldsymbol{A}^{-1}\left(\hat{\boldsymbol{\mu}} - r\boldsymbol{P} + \boldsymbol{\Lambda}\boldsymbol{\theta}^{\mathbf{0}}\right) \\ &\quad -\frac{1}{2}\left(\boldsymbol{A}^{-1}\left(\hat{\boldsymbol{\mu}} - r\boldsymbol{P} + \boldsymbol{\Lambda}\boldsymbol{\theta}^{\mathbf{0}}\right) - \boldsymbol{\theta}^{\mathbf{0}}\right)' \boldsymbol{\Lambda}\left(\boldsymbol{A}^{-1}\left(\hat{\boldsymbol{\mu}} - r\boldsymbol{P} + \boldsymbol{\Lambda}\boldsymbol{\theta}^{\mathbf{0}}\right) - \boldsymbol{\theta}^{\mathbf{0}}\right) \\ &= \frac{1}{2}\left(\hat{\boldsymbol{\mu}} - r\boldsymbol{P}\right)' \boldsymbol{A}^{-1}\left(\hat{\boldsymbol{\mu}} - r\boldsymbol{P}\right) + (\boldsymbol{\theta}^{\mathbf{0}})'\boldsymbol{\Lambda}'\boldsymbol{A}^{-1}\left(\hat{\boldsymbol{\mu}} - r\boldsymbol{P}\right) + \frac{1}{2}(\boldsymbol{\theta}^{\mathbf{0}})'(\boldsymbol{\Lambda}\boldsymbol{A}^{-1}\boldsymbol{\Lambda} - \boldsymbol{\Lambda})\boldsymbol{\theta}^{\mathbf{0}} . \end{aligned} \tag{4.40}$$

Given that $E_1(\hat{\boldsymbol{\mu}} - r\boldsymbol{P}) = \boldsymbol{\mu} - r\boldsymbol{P}$ and $V_1(\hat{\boldsymbol{\mu}} - r\boldsymbol{P}) = \boldsymbol{\Sigma} - \hat{\boldsymbol{\Sigma}}$, for any diagonal matrix $\boldsymbol{B}$, the expectation of $\left(\hat{\boldsymbol{\mu}} - r\boldsymbol{P}\right)' \boldsymbol{B}\left(\hat{\boldsymbol{\mu}} - r\boldsymbol{P}\right)$ is given by:

$$E_1\left(\left(\hat{\boldsymbol{\mu}} - r\boldsymbol{P}\right)' \boldsymbol{B}\left(\hat{\boldsymbol{\mu}} - r\boldsymbol{P}\right)\right) = \mathrm{Tr}(\boldsymbol{B}(\boldsymbol{\Sigma} - \hat{\boldsymbol{\Sigma}})) + (\boldsymbol{\mu} - r\boldsymbol{P})'\boldsymbol{B}(\boldsymbol{\mu} - r\boldsymbol{P}) .$$

Using the fact that $\boldsymbol{A}$ is a diagonal matrix, the expectation of U as given in equation (4.40) is given by:

$$\begin{aligned}
U_1 &= \frac{1}{2}\mathrm{Tr}\left(\boldsymbol{A}^{-1}(\boldsymbol{\Sigma}-\hat{\boldsymbol{\Sigma}})\right)+\frac{1}{2}(\boldsymbol{\mu}-r\boldsymbol{P})'\boldsymbol{A}^{-1}(\boldsymbol{\mu}-r\boldsymbol{P})+(\boldsymbol{\theta}^{\mathbf{0}})'\boldsymbol{\Lambda}'\boldsymbol{A}^{-1}(\boldsymbol{\mu}-r\boldsymbol{P})+\frac{1}{2}(\boldsymbol{\theta}^{\mathbf{0}})'(\boldsymbol{\Lambda}\boldsymbol{A}^{-1}\boldsymbol{\Lambda}-\boldsymbol{\Lambda})\boldsymbol{\theta}^{\mathbf{0}} \\
&= \frac{1}{2}\sum_{i=1}^{n}\frac{\Sigma_{ii}-\hat{\Sigma}_{ii}}{\gamma\hat{\Sigma}_{ii}+\Lambda_{ii}}+\frac{1}{2}\sum_{i=1}^{n}\frac{(\mu_i-rP_i)^2}{\gamma\hat{\Sigma}_{ii}+\Lambda_{ii}}+\sum_{i=1}^{n}\frac{(\mu_i-rP_i)\theta_i^0\Lambda_{ii}}{\gamma\hat{\Sigma}_{ii}+\Lambda_{ii}}+\frac{1}{2}\sum_{i=1}^{n}\frac{(\theta_i^0)^2\Lambda_{ii}^2}{\gamma\hat{\Sigma}_{ii}+\Lambda_{ii}}+\frac{1}{2}\sum_{i=1}^{n}(\theta_i^0)^2\Lambda_{ii}^2 \\
&= -\frac{1}{2\gamma}+\frac{1}{2}\sum_{i=1}^{n}\frac{\frac{1}{\gamma}\Lambda_{ii}+\Sigma_{ii}+(\mu_i-rP_i)^2+2(\mu_i-rP_i)\theta_i^0\Lambda_{ii}+(\theta_i^0)^2\Lambda_{ii}^2}{\gamma\hat{\Sigma}_{ii}+\Lambda_{ii}}+\frac{1}{2}\sum_{i=1}^{n}(\theta_i^0)^2\Lambda_{ii}^2\,.
\end{aligned}$$

Introducing the notation $a_i = \frac{1}{2\gamma}\Lambda_{ii}+\frac{1}{2}\Sigma_{ii}+\frac{1}{2}(\mu_i-rP_i+\theta_i^0\Lambda_{ii})^2$ and ignoring the constants, maximizing U_1 is equivalent to maximizing the expression

$$\sum_{i=1}^{n}\frac{a_i}{\gamma\hat{\Sigma}_{ii}+\Lambda_{ii}}\,.$$

	Additive learning technology	Entropy learning technology
Without transaction costs	*Specialized learning about the asset with the highest* $(\mu_i - P_i r)^2 + \Sigma_{ii}$	*Specialized learning about the asset with the highest* $(\mu_i - P_i r)^2 \Sigma_{ii}^{-1}$
Proportional transaction costs	*Generalized learning with no closed-form solution*	*Specialized learning with no closed-form solution*
Quadratic transaction costs	*Generalized learning and the optimal posterior variance solves* $(\frac{1}{\hat{\Sigma}_{ii}})^* = \max(\mathbf{\Sigma}_{ii}^{-1}, h_i^{-1}(\lambda_0))$	*Specialized or generalized learning* *Specialized learning about the asset with the highest* $-\frac{a_j}{\gamma\Sigma_{jj}+\Lambda_{jj}} + \frac{a_j}{\frac{\gamma\Sigma_{jj}}{e^K}+\Lambda_{jj}}$ *Generalized learning and the optimal posterior variance solves* $\ln\left(\frac{\Sigma_{ii}}{\hat{\Sigma}_{ii}}\right)^* = \max(0, g_i^{-1}(\lambda_0))$

TABLE 4.1: *Investors' information acquisition for different transaction costs and learning technologies.*

This table summarizes investors' information acquisition for different cases. We consider both proportional and quadratic transaction costs and compare the results with the case without transaction costs. Moreover, we consider both additive and entropy learning technologies. In the case with quadratic transaction costs and an entropy learning technology, whether investors choose specialized or generalized learning depends on the parameter values.

REFERENCES

Amin, K., J. D. Coval, and H. N. Seyhun, 2004, Index option prices and stock market momentum, *Journal of Business* 77, 835–874.

Ang, A., and G. Bekaert, 2002, International asset allocation with regime shifts, *Review of Financial Studies* 15, 1137–1187.

Ang, A., and J. Chen, 2002, Asymmetric correlations of equity portfolios, *Journal of Financial Economics* 63, 443–494.

Bachmann, R., S. Elstner, and E. R. Sims, 2010, Uncertainty and economic activity: Evidence from business survey data, *NBER Working Paper*.

Baker, M., and J. Wurgler, 2006, Investor sentiment and the cross-section of stock returns, *Journal of Finance* 4, 1645–1680.

Bakshi, G., N. Kapadia, and D. Madan, 2003, Stock returns characteristics, skew laws, and the differential pricing of individual equity options, *Review of Financial Studies* 16, 101–143.

Barberis, N., A. Shleifer, and J. Wurgler, 2005, Comovement, *Journal of Financial Economics* 75, 283–317.

Ben-David, I., F. Franzoni, and R. Moussawi, 2012, ETFs, arbitrage, and contagion, *Working paper*.

Bergman, Y., 1995, Option pricing with differential interest rates, *Review of Financial Studies* 53, 475–500.

Berkovich, E., and Y. Shachmurove, 2013, An error of collateral: Why selling SPX put options may not be profitable, *The Journal of Derivatives* 20, 31–42.

Bernardo, A. E., and O. Ledoit, 2000, Gain, loss and asset pricing, *Journal of Political Economy* 108, 144–172.

Bessembinder, H., and P. Seguin, 1992, Futures-trading activity and stock price volatility, *Journal of Finance* 47, 2015–2034.

Bloom, N., 2009, The impact of uncertainty shocks, *Econometrica* 77, 623–685.

Bollen, N. P. B., and E. R. Whaley, 2004, Does net buying pressure affect the shape of implied volatility functions?, *Journal of Finance* 59, 711–753.

Boudoukh, J., and M. Richardson, 1993, Stock returns and inflation: A long-horizon perspective, *American Economic Review* 83, 1346–1355.

Brenner, M., and D. Galai, 1996, Implied interest rates, *Journal of Business* 59, 493–509.

Campbell, J., M. Lettau, B. Malkiel, and Y. Xu, 2001, Have individual stocks become more volatile? An empirical exploration of idiosyncratic risk, *Journal of Finance* 56, 1–43.

Cochrane, J. H., and J. Saa-Requejo, 2000, Beyond arbitrage: Good-deal asset price bounds in incomplete markets, *Journal of Political Economy* 108, 79–119.

Constantinides, G. M., J. C. Jackwerth, and S. Perrakis, 2009, Mispricing of S&P 500 index options, *Review of Financial Studies* 22, 1247–1277.

Cvitanic, J., H. Pham, and N. Touzi, 1998, A closed form solution to the problem of super-replication under transaction costs, *Finance and Stochastics* 3, 35–54.

Cvitanic, J., H. Pham, and N. Touzi, 1999, Super-replication in stochastic volatility models under portfolio constraints, *Journal of Applied Probability* 36, 523–545.

Da, Z., and S. Shive, 2012, When the bellwether dances to noise: Evidence from exchange-traded funds, *Working paper.*

Domowitz, I., and B. Steil, 1999, Automation, trading costs and the structure of the securities trading industry, *Brookings-Wharton Papers on Financial Services.*

Driessen, J., P. Maenhout, and G. Vilkov, 2009, The price of correlation risk: Evidence from equity options, *Journal of Finance* 64, 5–59.

Dumas, B., and E. Luciano, 1991, An exact solution to a dynamic portfolio choice problem under transactions costs, *Journal of Finance* 46, 577–595.

Dybvig, P. H., 2005, Mean-variance portfolio rebalancing with transaction costs, *Working paper.*

Edwards, F. R., 1988, Futures trading and cash market volatility: Stock index and interest rate futures, *Journal of Futures Markets* 8, 421–439.

Erb, C., C. Harvey, and T. Viskanta, 1994, Forecasting international equity correlations, *Financial Analysts Journal* 50, 32–45.

Fremault, A., 1991, Stock index futures and index arbitrage in a rational expectations model, *Journal of Business* 64, 523–547.

French, K. R., G. W. Schwert, and R. F. Stambaugh, 1987, Expected stock returns and volatility, *Journal of Financial Economics* 19, 3–29.

Garleanu, N., and L. H. Pedersen, 2013, Dynamic trading with predictable returns and transaction costs, *Journal of Finance* 68, 2309–2340.

Garleanu, N., L. H. Pedersen, and A. Poteshman, 2009, Demand-based option pricing, *Review of Financial Studies* 22, 4259–4299.

Goetzmann, W., L. Li, and G. Rouwenhorst, 2005, Long term global market correlations, *Journal of Business* 78, 1–38.

Greenwood, R., 2005, Short- and long-term demand curves for stocks: Theory and evidence on the dynamics of arbitrage, *Journal of Financial Economics* 75, 607–649.

Grossman, S., and J. Stiglitz, 1980, On the impossibility of informationally efficient markets, *American Economic Review* 70, 81–101.

Han, B., 2008, Investor sentiment and option prices, *Review of Financial Studies* 21, 387–414.

Heston, S., 1993, A closed-form solution for options with stochastic volatility with applications to bond and currency options, *Review of Financial Studies* 6, 327–344.

Jackwerth, J. C., and M. Rubinstein, 1996, Recovering probability distributions from option prices, *Journal of Finance* 51, 1611–1631.

Jaffe, J. F., and G. Mandelker, 1976, The 'Fisher effect' for risky assets: An empirical investigation, *Journal of Finance* 31, 447–458.

Johannes, M., and S. Sundaresan, 2007, Pricing collateralized swaps, *Journal of Finance* 62, 383–410.

Jones, P. M., and E. Olson, 2013, The time-varying correlation between uncertainty, output, and inflation: Evidence from a DCC-GARCH model, *Economics Letters* 118, 33–37.

Kamara, A., and T. W. Miller, 1995, Daily and intradaily tests of European put-call parity, *Journal of Financial and Quantitative Analysis* 30, 519–539.

Krishnan, C., R. Petkova, and P. Ritchken, 2009, Correlation risk, *Journal of Empirical Finance* 16, 353–367.

Ledoit, O., P. Santa-Clara, and M. Wolf, 2003, Flexible multivariate GARCH modeling with an application to international stock markets, *Review of Economics and Statistics* 85, 735–747.

Levy, H., 1985, Upper and lower bounds of put and call option value: Stochastic dominance approach, *Journal of Finance* 40, 1197–1217.

Li, M., and N. Pearson, 2005, Price deviations of S&P 500 index options from the Black-Scholes formula follow a simple pattern, *AFA 2006 Boston Meetings Paper.*

Liu, H., 2004, Optimal consumption and investment with transaction costs and multiple risky assets, *Journal of Finance* 59, 289–338.

Liu, H., and M. Loewenstein, 2002, Optimal portfolio selection with transaction costs and finite horizons, *Review of Financial Studies* 15, 805–835.

Lo, A. W., H. Mamaysky, and J. Wang, 2004, Asset prices and trading volume under fixed transactions costs, *Journal of Political Economy* 112, 1054–1090.

Longin, F., and B. Solnik, 1995, Is the correlation in international equity market returns constant: 1960-1990?, *Journal of International Money and Finance* 14, 3–26.

Longin, F., and B. Solnik, 2001, Correlation structure of international equity markets during extremely volatile periods, *Journal of Finance* 56, 649–676.

Longstaff, F. A., 2001, Optimal portfolio choice and the valuation of illiquid securities, *Review of Financial Studies* 14, 407–431.

Longstaff, F. A., 2009, Portfolio claustrophobia: Asset pricing in markets with illiquid assets, *American Economic Review* 99, 1119–1144.

Lou, W., 2009, On asymmetric funding of swaps and derivatives — a funding cost explanation of negative swap spreads, *SSRN Working paper.*

Perrakis, S., and P. J. Ryan, 1984, Option pricing bounds in discrete time, *Journal of Finance* 39, 519–525.

Pindyck, R. S., and J. J. Rotemberg, 1993, The comovement of stock prices, *Quarterly Journal of Economics* 108, 1073–1104.

Piterbarg, V., 2010, Funding beyond discounting: Collateral agreements and derivatives pricing, *Risk* 2, 97–102.

Pollet, J. M., and M. Wilson, 2010, Average correlation and stock market returns, *Journal of Financial Economics* 96, 364–380.

Ritchken, P., 1985, On option pricing bounds, *Journal of Finance* 39, 519–525.

Santa-Clara, P., and A. Saretto, 2009, Option strategies: Good deals and margin calls, *Journal of Financial Markets* 12, 391–417.

Schürhoff, N., and A. Ziegler, 2011, Variance risk, financial intermediation, and the cross-section of expected option returns, *CEPR Discussion Paper No. DP8268.*

Staer, A., 2012, Equivalent volume and comovement, *Working paper*.

Van Nieuwerburgh, S., and L. Veldkamp, 2009, Information immobility and the home bias puzzle, *Journal of Finance* 64, 1187–1215.

Van Nieuwerburgh, S., and L. Veldkamp, 2010, Information acquisition and under-diversification, *Review of Economic Studies* 77, 779–805.

Verrecchia, R. E., 1982, Information acquisition in a noisy rational expectations economy, *Econometrica* 50, 1415–1430.

Vijh, A., 1994, S&P 500 trading strategies and stock betas, *Review of Financial Studies* 75, 215–251.

Appendices

Lujing Su

Nationality: Chinese
Date of Birth: 06.06.1985

CURRENT EMPLOYER	Senior Quantitative Analyst Deutsche Bank Risk Center, Berlin -Validating equity derivative pricing models	2015-present
EDUCATION	*Ph.D. Candidate,* Swiss Finance Institute and University of Zurich Research Interests: Derivative Pricing, Asset Allocation, Risk Management Dissertation Committee: Prof. Dr. Markus Leippold, Prof. Dr. Alexandre Ziegler	2009-2015
	M.Sc. Mathematical Economics, Rank 1 out of 40, Bielefeld University, Germany	2008-2009
	M.Sc. Mathématiques Appliquées en Economie et en Finance Rank 1 out of 40, University of Paris 1, France	2007-2008
	B.Sc. in Mathematics and Applied Mathematics (with honors), Zhejiang University, China	2003-2007
TEACHING ASSISTANCE	Dynamic Portfolio Theory and Asset Pricing (PhD course), Zurich	2010-2014
INTERNSHIPS	Industrial and Commercial Bank of China, Hangzhou, China -Assisted wealth manager in informing clients about financial products -Provided advice on customization of structures to best fit their needs	2008